Atmospheric and Environmental Research

131 Hartwell Avenue
Lexington, Massachusetts
02421-3126
USA
Tel: +1 781 761-2288
Fax: +1 781 761-2299
www.aer.com

Final Report
Revision 2.0

Source Apportionment of Fine Particulate Matter at Big Bend National Park

Task 2 – PMF Analysis of Big Bend National Park IMPROVE Aerosol Data

TCEQ Contract No. 582-15-50414
Work Order No. 582-16-62291-02

Prepared by:
Jennifer Hegarty, Rebecca Adams-Selin and Matthew Alvarado
Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Ave.
Lexington, MA 02421
Correspondence to: jhegarty@aer.com

Prepared for:

Erik Gribbin
Texas Commission on Environmental Quality
Air Quality Division
Building E, Room 342S
Austin, TX 78711-3087

June 30, 2016

Document Change Record

Revision	Revision Date	Remarks
0.1	**08 June 2016**	**Internal draft for review**
1.0	**10 June 2016**	**Draft Version of Final Report Submitted to TCEQ as Deliverable 3.1**
2.0	**30 June 2016**	**Final Report Submitted to TCEQ as Deliverable 3.2**

Table of Contents

List of Acronyms

As ..... Arsenic
ASOS ..... Automated Surface Observing System
BIBE ..... Big Bend National Park
BRAVO..... Big Bend Regional Aerosol and Visibility Observational study
Ca ..... calcium
Cl..... chlorine
COHA ..... Causes of Haze Assessment
EC ..... elemental carbon
EGU ..... Electrical Generating Units
Fe..... iron
GUI ..... Graphical User Interface
K..... potassium
IMPROVE..... Interagency Monitoring of Protected Visual Environments
ME2..... Multilinear Engine version 2
METAR..... Meteorological Terminal Aviation Routine weather report
Mg ..... magnesium
Na ..... sodium
NO3..... nitrate
OC ..... organic carbon
Pb ..... lead
PM2.5..... Fine (<2.5 microns diameter) Particulate Matter
PMF..... Positive Matrix Factorization
Si ..... silicon
STILT..... Stochastic Time-Inverted Lagrangian Transport
Zn ..... Zinc

List of Figures

List of Tables

Executive Summary

The purpose of this project was to perform a source apportionment of fine particulate matter ($PM_{2.5}$) in Big Bend National Park (BIBE) for the period of 2011 – 2014 in order to update TCEQ's knowledge regarding the sources of fine particulate matter impacting visibility in the park. We used daily average concentrations of speciated aerosols measured in BIBE from the IMPROVE monitoring network. The first step was to perform a factorization analysis on the aerosol measurements using the EPA Positive Matrix Factorization (PMF) tool to identify the major species source contributions to the measured $PM_{2.5}$ concentrations. The baseline PMF analysis, run from a graphical user interface (GUI), identified 6 source contribution factors to be biomass burning smoke, heavy metal dust, secondary sulfate pollution, a sea salt nitrate mixture, mineral dust, and secondary nitrate pollution. Smoke and mineral dust source contribution factors were episodic while other factors, including secondary sulfate and nitrate pollution, seemed more consistent over time. Sulfate was the largest contributing factor accounting for 42.7 % followed by mineral dust at 24.1 % and biomass burning smoke at 17.4 %. Nitrate pollution was a relatively minor factor contributing no more than 12.9 % but occurred primarily in a mixture with sea salt and was therefore likely less. For the days with the highest 20 % of $PM_{2.5}$ measurements the secondary sulfate pollution contribution decreased to 32.4 % while contributions from likely natural sources such as mineral dust and biomass burning smoke increased to 31.0 % and 21.0 %, respectively. The opposite trend was found for the days with the lowest 20 % of $PM_{2.5}$ measurements as the sulfate contribution was 48.7 % while the mineral dust contribution was 19.0 % and smoke was 18.9 %.

The baseline PMF results were compared with an analysis of the period from 2000 – 2004 conducted as part of the Causes of Haze Assessment study (COHA) to assess changes in the source apportionment over time. The newer results were in general agreement with the earlier results indicating that sulfate, dust, and smoke were the major contributing factors to $PM_{2.5}$ concentrations, though in the earlier assessment the sulfate and dust contributions were about equal. However, an important difference was that the proportional contribution distribution from the earlier period showed no sensitivity to the $PM_{2.5}$ concentrations.

The baseline PMF results were also evaluated with respect to their correlation with meteorological and time factors. The days with the highest 20 % of $PM_{2.5}$ were warmer and more humid than the days with the lowest 20 % of $PM_{2.5}$. The sulfate, smoke, and sea salt / nitrate factors were most strongly correlated with warmer temperatures and higher dewpoints, particularly in the winter and spring seasons. An important result was that local wind directions and wind speeds were not highly correlated with any of the source factor contributions. A slight relationship was evident on high $PM_{2.5}$ days in the fall and winter, with winds more frequently from the north than on low $PM_{2.5}$ days. However, the factor contributions showed seasonal variability, with smoke peaking in spring and dust in the fall, while sulfate pollution seemed to peak on Tuesdays and Saturdays, possibly indicating a time lag due to transport time from distant emissions sources.

We attempted to incorporate meteorological variables into the baseline PMF model as run from the GUI, but found that it was only capable of handling concentration data. Therefore, we utilized an enhanced version of the PMF called the Expanded Parametric Model that was included as part of the EPA PMF software package. We found that the expanded parametric model is a reasonable way to add additional meteorological and other data (specifically, wind speed, wind direction, season, and day of week) to the PMF analysis. In our analysis, the best results were obtained when

performing a 7-factor fit; however, the contribution estimates from the 7-factor expanded parametric model are not significantly different from our original baseline 6-factor PMF fit. In addition, using the expanded parametric model requires performing all of the analysis outside of the GUI of the EPA PMF software, and so could be more labor intensive. Thus we would recommend using the expanded parametric model only for cases where prior knowledge of the sources suggests that there should be a significant dependence of specific sources on wind direction, such as when two sources with similar species profiles are located near the measurement location but at different directions. A remote location like BIBE is thus probably not the best location for the use of the expanded parametric model.

Finally, we investigated the potential source regions using footprints generated with the STILT model. Our footprint analysis indicated that the strongest source region for all factors was from south of and west of BIBE in north central Mexico. However, strong source contributions were also possible for sulfate, smoke, dust, and nitrate from central Texas and for sulfate pollution and smoke from eastern Texas.

1 Introduction

1.1 Project Objectives

This project used the EPA Positive Matrix Factorization tool (EPA PMF v5.0, *Norris et al., 2014*) to conduct a detailed aerosol source apportionment analysis for fine particulate matter ($PM_{2.5}$) pollutants measured at Big Bend National Park (BIBE).

The objectives of the project were to:

- Provide source apportionment of aerosol data collected at BIBE on days with the highest twenty (20) percent 24-hour $PM_{2.5}$ measurements,
- Provide source apportionment of aerosol data collected at BIBE on days with the lowest twenty (20) percent 24-hour $PM_{2.5}$ average measurements, and
- Determine the technical feasibility of including relevant seasonal and meteorological variables into the PMF aerosol source apportionment analysis.

1.2 Purpose and Background

The purpose of this Work Order was to update the TCEQ's state of knowledge regarding the sources of $PM_{2.5}$ species measured at Big Bend National Park (BIBE). This analysis relied on state-of-the-art source apportionment techniques and speciated $PM_{2.5}$ data collected at the park through the Interagency Monitoring of Protected Visual Environments (IMPROVE) monitoring program. The last known source apportionment study of Big Bend was conducted as part of the Causes of Haze Assessment (COHA) performed by the Desert Research Institute using $PM_{2.5}$ data from the 2001-2004 time frame *[Green, 2006]*. Prior to that analysis, the Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study field campaign was conducted from July through October 1999 *[Pitchford et al., 2004]* and a source apportionment was performed. Since 2004, air quality strategies (e.g. substitution of low-sulfur coal for lignite coal at coal-fired Electrical Generating Units (EGUs) or higher use of natural gas powered EGUs) may have changed the composition of visibility reducing pollutants in important ways. Updated source apportionment information is essential to understand any changes in the composition or seasonality of $PM_{2.5}$ pollutants or identify changes in source areas. Therefore, for this study we analyzed more recent IMPROVE aerosol data from BIBE collected during 2011 – 2014. The schedule for the project is given in Table 1.

1.3 Report Outline

This Final Report documents the methods and pertinent accomplishments of this project, including comprehensive overviews of each task, a summary of the data collected and analyzed during this work, key findings, shortfalls, limitations and recommended future tasks. It satisfies Deliverable 3.2 of the Work Plan for Work Order No. 582-16-62291-02.

Deliverable 3.2: Final Report for TCEQ review and approval, delivered electronically via file transfer protocol or e-mail in Microsoft Word format and PDF format

Deliverable Due Date: June 30, 2016

We first performed a baseline PMF analysis with only the speciated $PM_{2.5}$ data (Section 2). As part of this work, we examined the relationship between meteorological and time factors (seasonal and day-of-the-week) and the source contribution factors identified in the baseline PMF analysis (Section 2.2). We then investigated the feasibility of incorporating other types of data such as meteorological variables into this baseline analysis both directly (Section 2.4) and using a more

rigorous approach to incorporate meteorological, seasonal, and day-of-the-week variables into an expanded parametric model (Section 3) that is included in the EPA PMF software package but not fully integrated into the graphical user interface (GUI). Finally, we examined footprints generated by the STILT model *[Lin et al., 2003]* to examine the role of transport and identify the likely locations of the various sources contributing to $PM_{2.5}$ concentrations in BIBE (Section 4).

Table 1. **Projected Schedule for TCEQ Work Order No. 582-16-62291-02**

Milestones	Planned Date
Task 1 - Work Plan	
1.1: TCEQ-approved Work Plan	Feb. 8, 2016
1.2: TCEQ-approved QAPP	Feb. 8, 2016
Task 2 - PMF Analysis of Big Bend National Park IMPROVE Aerosol Data	
2.1: A short technical memorandum on the integration of seasonal and meteorological variables in the PMF analyses.	May 31, 2016
Task 3 - Draft and Final Reports	
3.1: Draft Report	June 10, 2016
3.2: Final Report	June 30, 2016
3.3: Data files used in the PMF Analyses	June 30, 2016

2 Baseline PMF

The baseline source apportionment of $PM_{2.5}$ at BIBE was performed using Version 5.0 of the EPA PMF model *[Norris et al., 2014]*. The PMF software was installed on PCs at AER running the Windows 7 operating system. This model is run from a graphical user interface (GUI) and requires two input files, one for speciated concentration measurements and the other for speciated measurement uncertainty. The files may be in either .csv or. txt format. From the GUI the user can select from a number of options such as which species is the "total variable" whose concentration all the component species are contributing to ($PM_{2.5}$ in our case), the species to include in the factorization, the influence of each species (i.e. some species can be designated as "weak" so they have little impact on the calculations) and how many factors to include. Users can also run tests for robustness and create plots.

2.1 Source Contribution Factors

We performed a six-factor PMF analysis of the BIBE IMPROVE dataset for the years 2011 – 2014. The IMPROVE dataset includes daily average concentrations and measurement uncertainty estimates of speciated aerosols measured in BIBE. Before running the PMF the species list was edited to match as closely as possible that used for the COHA PMF modeling of the 2000 – 2004 period to facilitate comparisons between the datasets and identify any changes in the aerosol sources. The choice of 6 factors was also influenced by the COHA analysis which also identified six factors. We performed some sensitivity runs with more factors but did not find that it provided any additional meaningful information to our baseline analysis and so these results are not presented here. The six-factor PMF runs were also tested using the displacement option available in PMF 5.0 *[Norris et al., 2014]* and found to be robust.

The species profiles for the six factors for our baseline PMF runs are presented in Figure 1 to Figure 6. We have made the following identifications based on these profiles:

- Factor 1: The high levels of organic carbon (OC) and elemental carbon (EC), combined with the relatively high potassium (K) level, suggests that this factor is smoke from biomass burning, or "Smoke".
- Factor 2: This factor has high amounts of Arsenic (As) and lead (Pb) and Zinc (Zn) and may be heavy metal dust and will be referred to hereafter as "Dust (As, Pb)."
- Factor 3: Given the high amounts of sulfate we identify this factor as primarily secondary sulfate pollution or "Sulfate."
- Factor 4: This factor has high sodium (Na) and magnesium (Mg) and some chlorine (Cl) content, suggesting that it is from sea salt. It also has some nitrate (NO_3). Therefore, we identify this factor as a mixture of sea salt and nitrate or "Sea Salt / Nitrate."
- Factor 5: This factor has high amounts of calcium (Ca), iron (Fe) and silicon (Si) and so appears to be natural mineral dust, or "Dust (Ca, Fe, Si)."
- Factor 6: This factor includes trace amounts of nitrate (NO_3). Thus we call this factor as "Nitrate".

Our baseline PMF analysis indicated that smoke and mineral dust source contribution factors were episodic while other factors, including secondary sulfate and nitrate pollution, seemed more

consistent over time (Figure 7). The analysis also indicated that sulfate secondary pollution accounted for the greatest proportion, 42.7 %, of the fine particle mass, while the combination of both dust types accounted for 27.0 % with mineral dust being the much greater contributor at 24.1 %, with smoke contributing 17.4 %, sea salt and nitrate-rich secondary aerosol mixture 11.9 %, and trace nitrate only 1.0 % (Figure 8). For the days with the highest 20 % of $PM_{2.5}$ concentrations the sulfate-rich secondary aerosol contribution decreased by almost a quarter to 32.7 % while the dust combination increased to 32.9 % and became the major contributor (Figure 9). The smoke contribution also increased slightly to 21.0 % while the sea salt / nitrate and mainly nitrate factors saw minor changes to 13.0 % and 0.5 % respectively. In contrast for the days with the lowest 20 % of $PM_{2.5}$ concentrations the sulfate-rich secondary aerosols were by far the greatest contributor at 48.7 % followed by the dust combination (23.9 %), smoke (18.9 %), the sea salt /nitrate combination (6.0 %) and trace nitrate (2.5 %) which was a five times increase from proportional contribution over the high $PM_{2.5}$ days' contribution (Figure 10). Note that the proportionality of dust types also changed with the proportional contribution from heavy metal dust (Dust As, Pb, Zn) being the lowest (1.9 %) when $PM_{2.5}$ concentrations were high (Figure 9) and the highest (4.9 %) when $PM_{2.5}$ concentrations were low (Figure 10). Overall comparing the days with the highest and lowest $PM_{2.5}$ concentrations it appears as though when $PM_{2.5}$ concentrations are high there is a greater proportional contribution from "natural sources" such as mineral dust and smoke (possibly from wildfires) and on days when the $PM_{2.5}$ concentrations are low there is a greater proportional contribution from anthropogenic pollution sources that create secondary sulfate and to a much lower degree nitrate aerosols (Nitrate).

Our factor identifications were generally similar to the COHA analysis *[Green 2006]*, which identified the six factors for the period of 2000 – 2004 as mixture, smoke, nitrate-rich secondary, sulfate-rich secondary, and two dust types. Our factor proportional contributions were also similar. When all days were considered the COHA analysis indicated that sulfate-rich secondary aerosols (37 %) and dust (38 %) were the major contributors followed by smoke (15 %) nitrate-rich secondary (9 %) and mixture (1 %) (Figure 8). In contrast, including only the days with the highest 20 % of $PM_{2.5}$ concentrations made only minor changes to the proportional factor contributions for the COHA study with only 1% increases in sulfate-rich secondary aerosol (38 %), dust (39 %), and nitrate-rich secondary aerosol (10 %) and a small 3 % decrease in smoke (12 %) with mixture unchanged (Figure 9). Overall the biggest change in relative aerosol contributions between the COHA period and our more recent study period was that days with high $PM_{2.5}$ in 2011 – 2014 are associated with greater proportional contributions from natural (dust and smoke) sources.

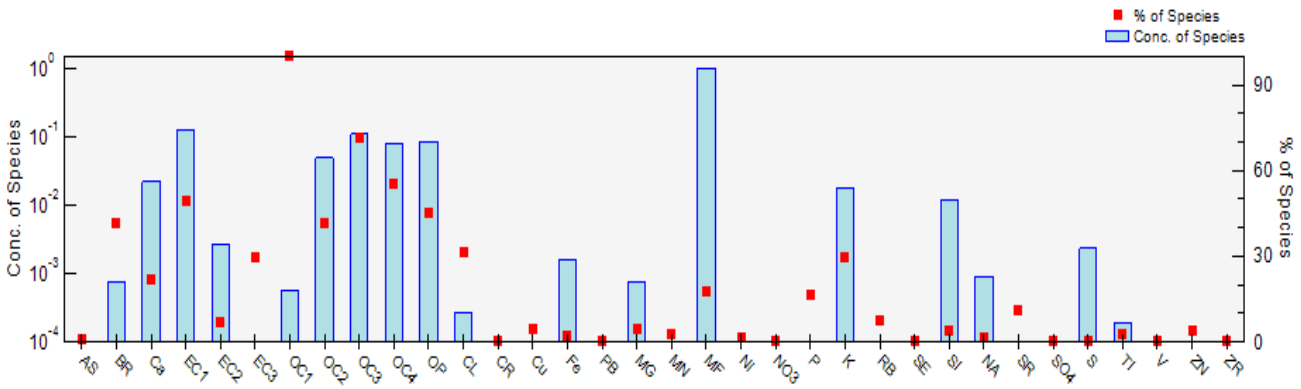

Figure 1. Species profile for factor 1, identified as a smoke from biomass burning ("Smoke"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

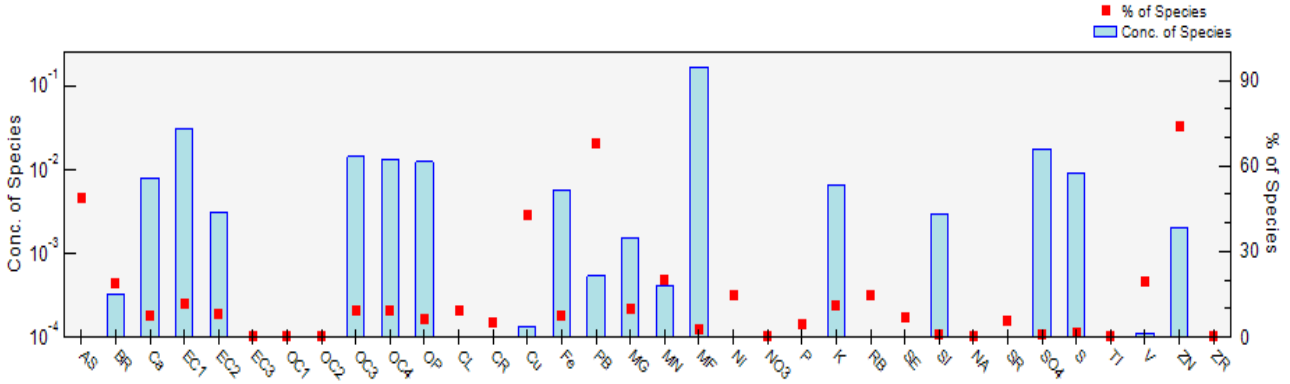

Figure 2. Species profile for factor 2, identified as heavy metal dust with arsenic and lead from biomass burning ("Dust (As, Pb)"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

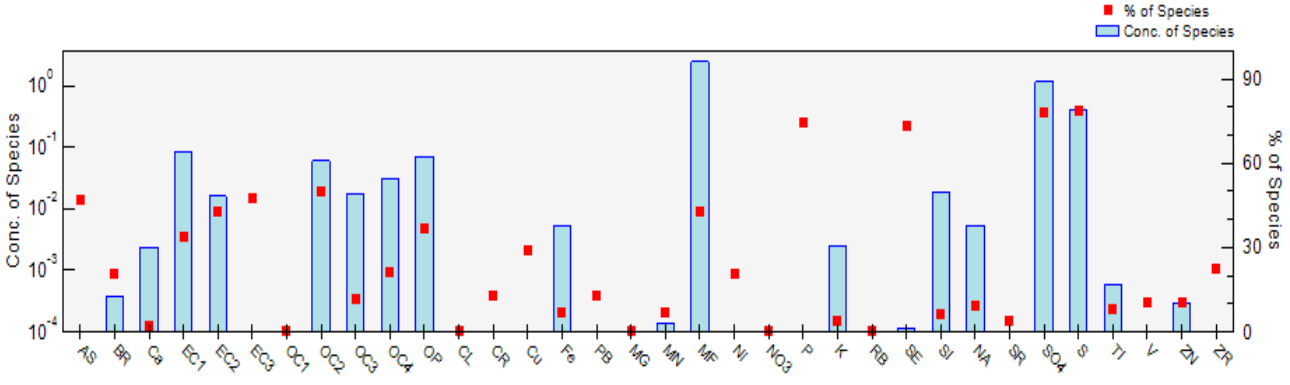

Figure 3. Species profile for factor 3, identified as secondary sulfate pollution ("Sulfate"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

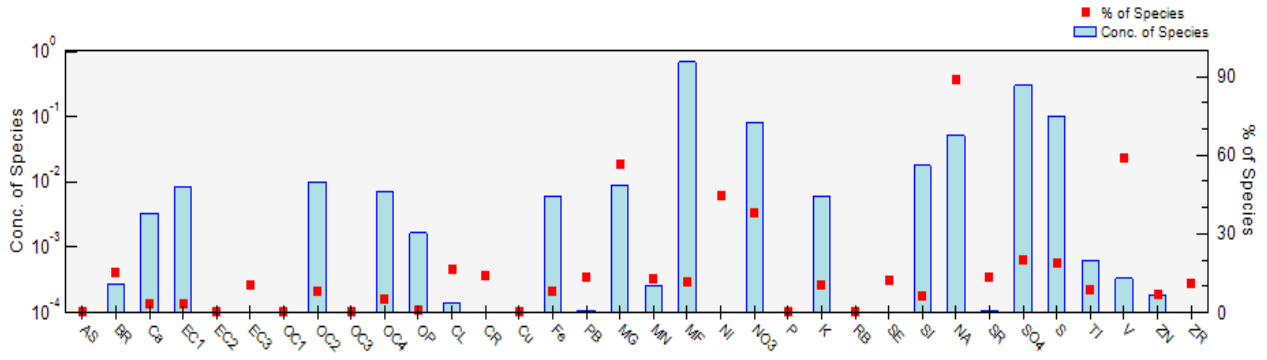

Figure 4. Species profile for factor 4, identified as a mixture of sea salt and nitrate ("Sea salt / Nitrate"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

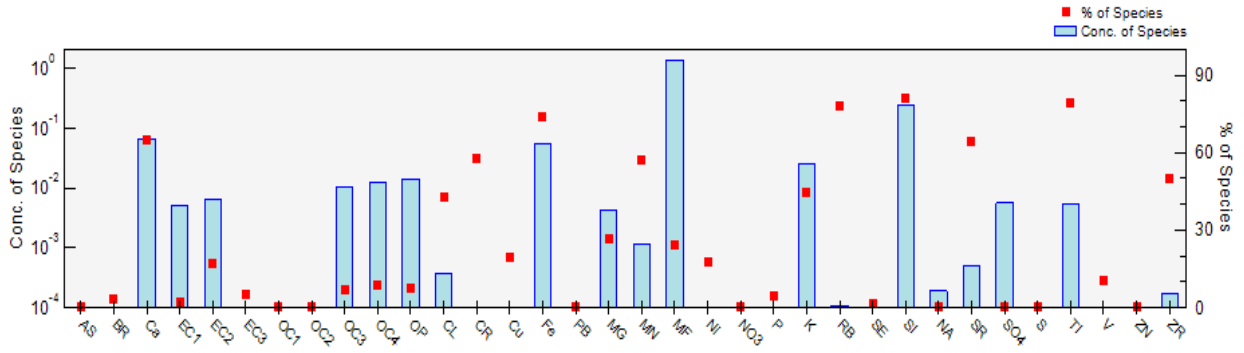

Figure 5. Species profile for factor 5, identified as a mineral dust ("Dust (Ca, Fe, Si)"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

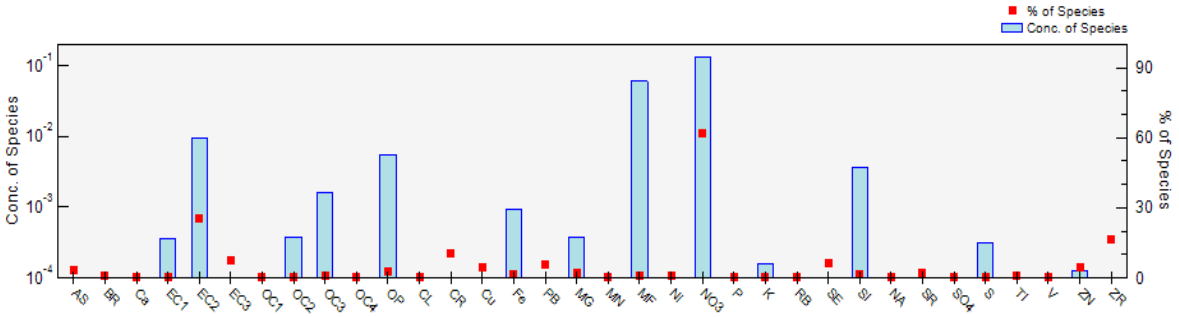

Figure 6. Species profile for factor 6, identified as secondary nitrate pollution ("Nitrate"), from the 6-factor baseline PMF model fit. Blue bars show the normalized concentration of the species (ug/m^3) in the factor, while the red dots show the percentage of that species in the factor.

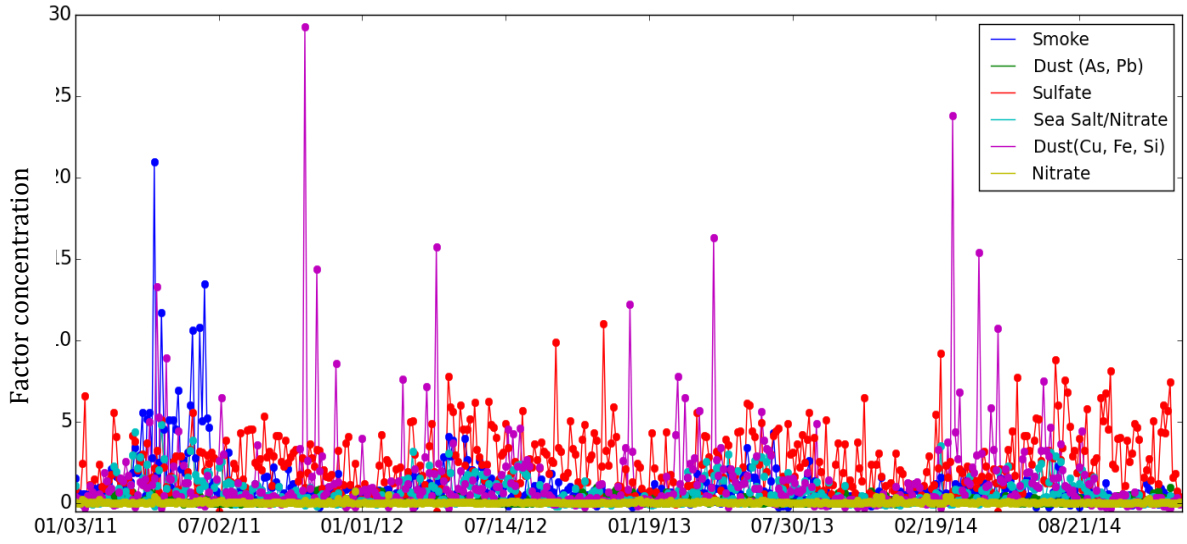

Figure 7. Time series of factor contributions in ug/m^3 to $PM_{2.5}$ measured concentrations at BIBE during 2011 – 2014 calculated from the 6-factor baseline PMF run.

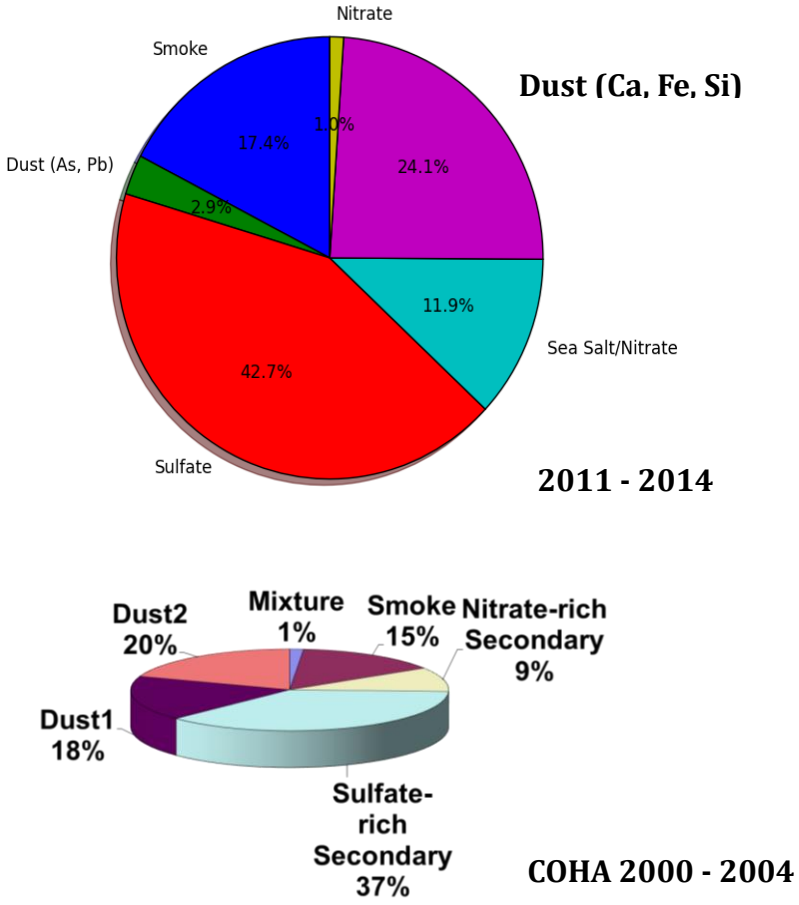

Figure 8. Pie chart of source factor contributions to the $PM_{2.5}$ concentrations for baseline PMF run for 2011 – 2014 (top) and COHA PMF runs for 2000 – 2004 (Bottom) for all days.

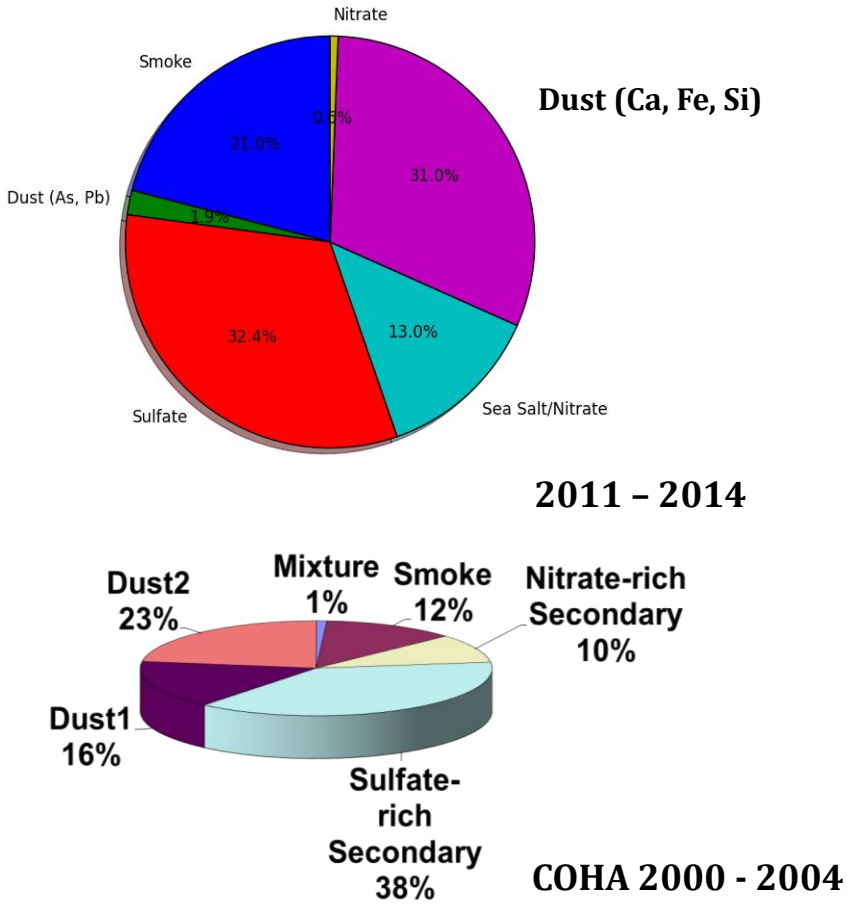

Figure 9. Same as Figure 8 but for days with the highest 20 % of $PM_{2.5}$ concentrations.

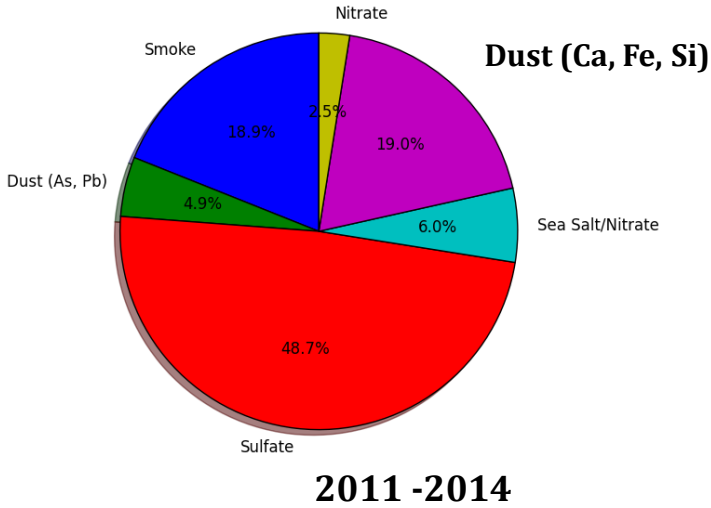

Figure 10. Pie chart of source factor contributions to the $PM_{2.5}$ concentrations for the baseline PMF run for 2011 – 2014 for days with the lowest 20 % of $PM_{2.5}$ concentrations.

2.2 Meteorological and Time Factor Analysis

We investigated whether there were patterns in local meteorological conditions associated with high contributions from each factor found by the baseline PMF analysis during each day-of-the-week and season. We used METAR (Meteorological Terminal Aviation Routine weather report) observations, available at 20 minute intervals, from a nearby Automated Surface Observing System (ASOS) station to characterize the meteorological conditions. We ran analyses using three different ASOS stations: KPRS (Presidio Lely, to the west of the park), K6R6 (Dryden, to the east of the park), and KE38 (Alpine-Casparis Airport, to the north of the park). Wind direction observations from KPRS and K6R6 were primarily from the north and east over the four-year period, and did not appear to match well with typical prevailing synoptic winds in the area. Potentially some local topographical features influence the wind directions observed at these stations. However, wind directions at KE38 (Alpine) observed over the four-year period were primarily from the south and west, and did fit with typical prevailing winds. Thus the rest of the analysis presented uses observations from KE38.

For each PMF factor, we first found all the dates when that factor's contribution was above the 60^{th} percentile of that distribution. We then calculated an average temperature and dewpoint, using observations from KE38, at each of those dates. Wind directions were divided into sixteen orthants (N, NNE, NE, ENE, E, etc.), following the method of *Chan et al. [2011]*, and the most frequent wind direction orthant for each day used. We then calculated the mean temperature and dewpoint, and most frequent wind direction, for the dates within each season, and for each day of the week.

A few interesting results were immediately apparent from this analysis. Figure 11 shows that there is not a significant signal in temperature or dewpoints with respect to days of the week. However, it is evident from these plots that the factor contributions from sulfate and sea salt/nitrate are higher on days with higher temperatures and significantly higher dewpoints. This is consistent with our conceptual model for those factors: sulfate is rapidly formed from SO_2 when clouds are present (high humidity) and sea salt should occur on days when moist sea air is reaching BIBE. Because the dewpoint signal is so strong, it is likely that the signal at higher temperatures is a result of the signal at higher dewpoints rather than representing independent information. The wind direction analysis (not shown) in general showed little difference in wind direction over factor, season, or day of week, instead only showing the prevailing southerly winds.

We repeated the analysis method described above, but over only the dates corresponding to the upper and lower 20^{th} percentiles of the observed range of $PM_{2.5}$. Figure 12 shows that the mean temperatures on the upper 20^{th} percentile $PM_{2.5}$ days were significantly warmer than the lower 20^{th} percentile days. This result is particularly noticeable in winter and spring, and for smoke throughout the year. This result for smoke is expected as warmer temperatures typically correspond with more favorable conditions for fire. Wind direction analyses for the upper and lower 20^{th} percentile days (not shown), when compared to the analyses for all factors, did reveal that in the fall on high $PM_{2.5}$ days the wind direction for all factors was more likely to be from the north instead of the south. This tendency was also evident in winter for the dust (Ca, Fe, Si) and nitrate factors.

The above analysis was repeated using the temperature, dewpoint, and wind direction at time of maximum heating during the day, instead of the daily average, but the results were not significantly different.

In summary, while the results we have found here using meteorological and seasonal analysis are consistent with our identification of factors we found with the PMF analysis, we did not find evidence that suggested that these could be more finely divided into additional factors. The particulate sources are some distance from the receptors, so this result is not entirely unexpected.

2.3 Mean Factor Contributions

Factor analysis was also performed by examining the mean factor contributions themselves on upper and lower 20th percentile days. Figure 13 displays the mean factor contributions for each season and day of week on the upper 20th percentile days. Mean factor contributions on the lower 20th percentile days were almost uniformly low and are not shown. From Figure 13 (top) it is evident that the factor contributions of Dust (As, Pb, Zn) and Nitrate were minimal throughout all seasons and days of week on high $PM_{2.5}$ days throughout all seasons and days of the week. Contributions from smoke peaked in the spring coinciding with the peak of the Texas and Mexico fire seasons, with the largest contributions coming from the spring 2011 wildfires. The factor contributions from Dust (Calcium, Iron, and Silicon) were much larger in magnitude than the Dust (As, Pb, Zn) contributions, and peaked in the fall and spring. The peak in fall dust contributions were due almost entirely to dust events in October 2011; otherwise the majority of high dust events occurred in the spring, corresponding well with drought and fire season. Sulfate contributions peaked in the fall and winter.

The bottom of Figure 13 shows the mean factor contributions by day of week. Sulfate contributions were largest on Tuesdays and Saturdays, potentially indicating a peak in fossil fuel combustion near those days. Because the particulate sources are some distance from the receptors, it is likely there is a lag between the actual combustion and the measurement of the sulfate in BIBE, so the peak fossil fuel combustion could have occurred on Monday and Friday. Factor concentrations from Dust (Calcium, Iron, and Silicon) had a significant peak on Wednesday and a lesser peak on Thursday, the reasons of which are not clear.

2.4 Incorporation of Meteorological Variables

We attempted to incorporate meteorological variables directly into the EPA PMF 5.0 tool *[Norris et al., 2014]* using the PMF GUI tool in the same way as the concentration data. For this feasibility test we included only winds observed at Alpine-Casparis Airport to the north of BIBE as these winds were most consistent with the prevailing winds of surrounding locations. The daily average wind speed in m s^{-1} and the most frequent wind direction in 16 orthants (N, NNE, NE, ENE, E, etc.) following the method of *Chan et al. [2011]* were used. Uncertainty estimates were set at 1 m s^{-1} and 10° for wind speed and direction respectively.

The PMF runs were performed with 6 and 8 factors. Both sets of runs converged within 20 iterations. The analysis did not indicate any strong correlations between the wind speed and direction and the source contribution factors. In addition, the source contribution factor magnitudes for each of the species were generally reduced as some of the magnitude was distributed to the wind speed and direction variables. Increasing the number of factors from 6 to 8 did not provide any additional information to analysis. The factor profiles for 6 of the factors in the 8-profile case corresponded to those in the 6-factor case but the source contribution species associated with the additional species could not be clearly identified (not shown). Overall these tests indicated that incorporating meteorological or other non-concentration variables into the PMF 5.0 program using the GUI tool will not produce worthwhile results. Therefore, as described in Section 3, we use an

enhanced PMF tool referred to as the Expanded Parametric Model specifically designed to handle other data types.

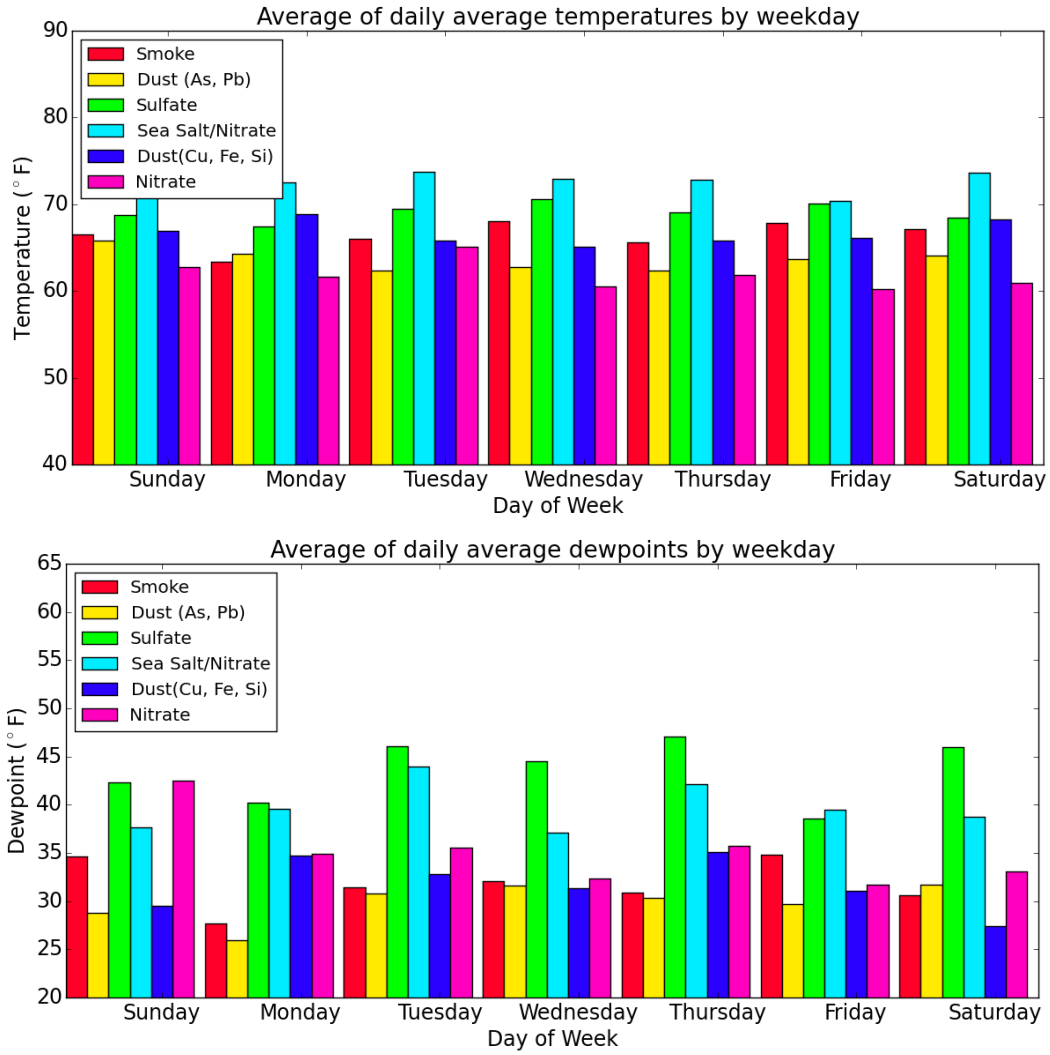

Figure 11. Mean daily average temperatures (top) and dewpoints (bottom) for each day of the week, for each factor.

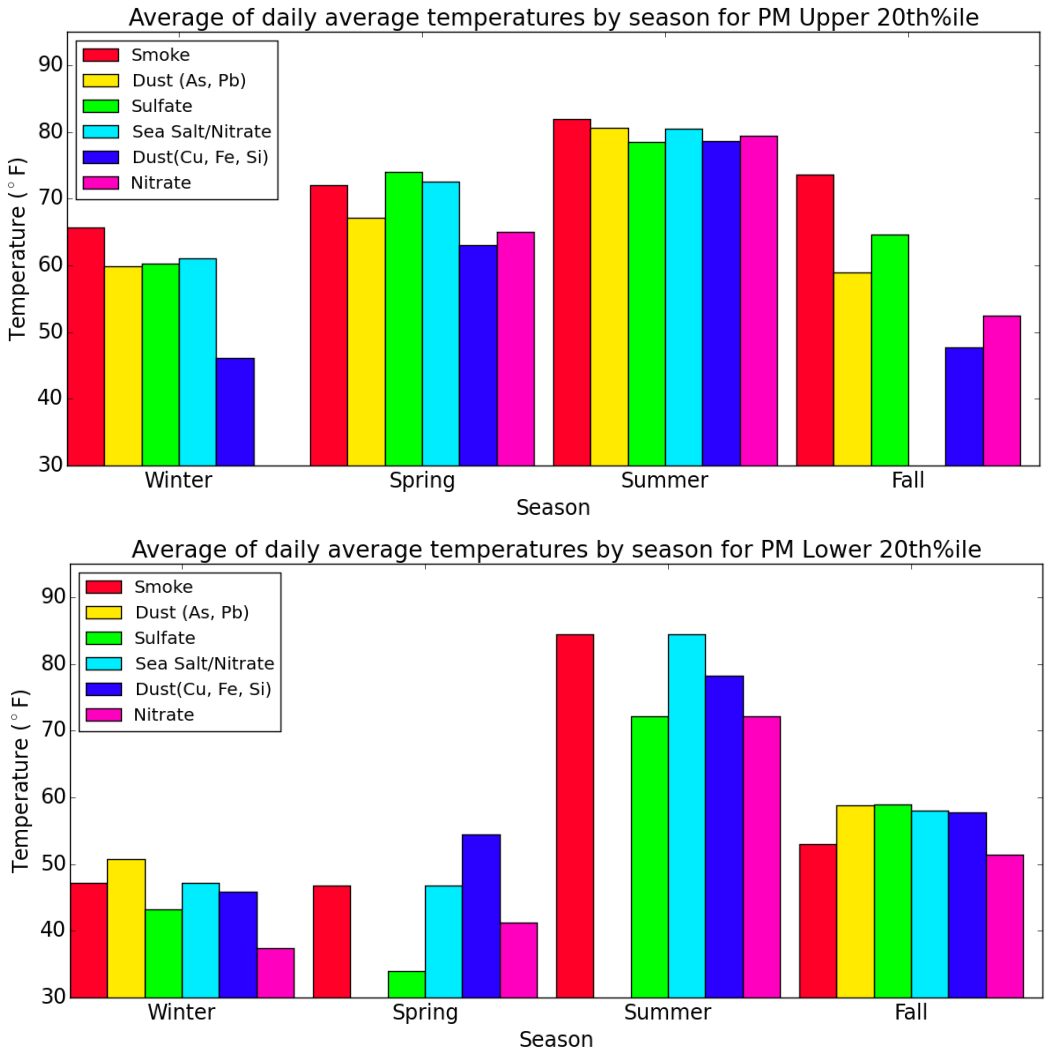

Figure 12. Mean daily average temperatures for each season for the upper (top) and lower (bottom) 20th percentile of $PM_{2.5}$. If a column is missing, it is because there were no dates above the 60th percentile for that factor contributions that were also in the upper (or lower) 20th percentile of $PM_{2.5}$ days.

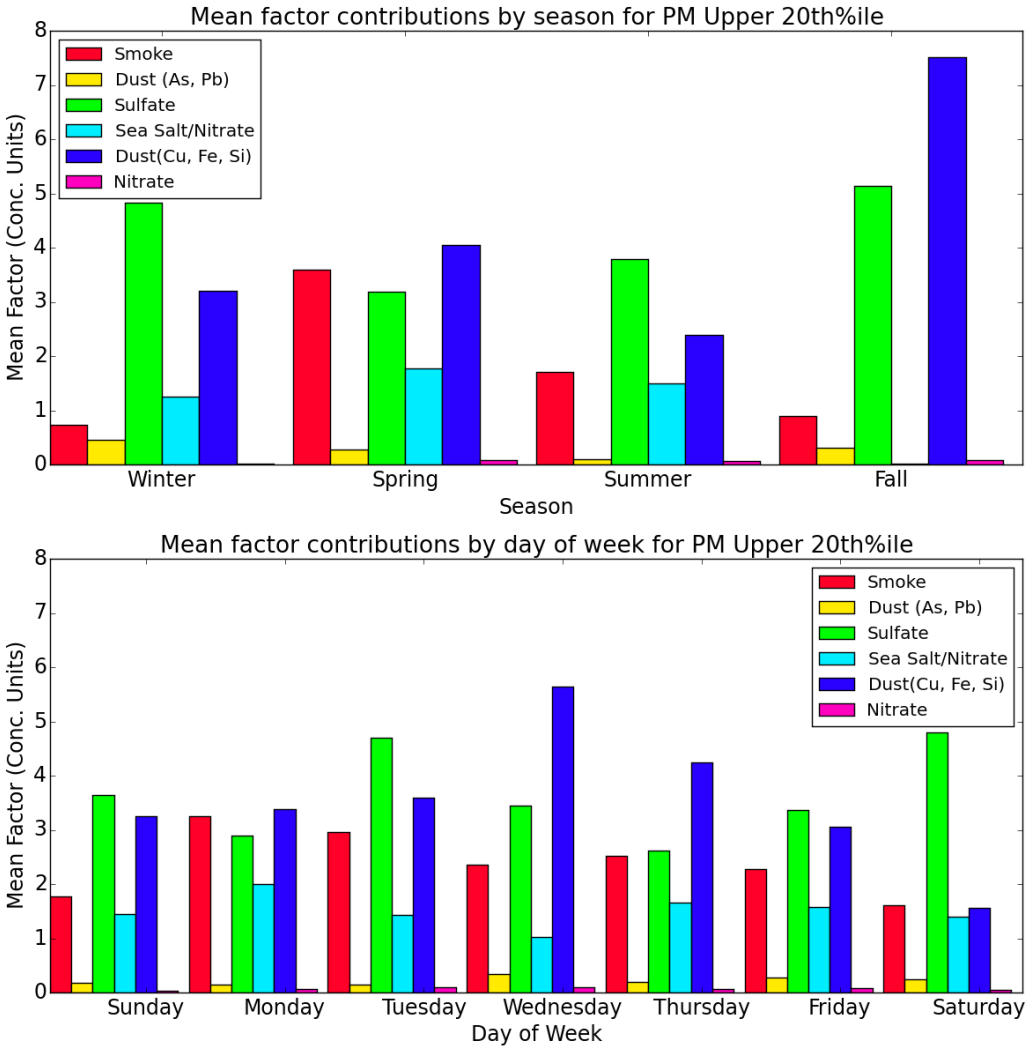

Figure 13. Mean factor contributions (concentration units) by season (top) and day of week (bottom) for days containing $PM_{2.5}$ concentrations in the upper 20th percentile.

3 Expanded Parametric Model

The most rigorous attempt to directly include meteorological variables into the PMF fitting procedure that we have been able to find is the expanded parametric model of *Paatero and Hopke [2002]*. In this approach, data on wind speed, wind direction, day of week, and season are included in the PMF analysis. However, as the dependence of the observed concentration on wind variables is expected to be highly non-linear, the wind data are not included in the PMF as dependent variables, like the different elemental species, but instead are included as independent variables *[Paatero and Hopke, 2002]*. This approach has been implemented as an option in the EPA PMF software *[Paatero, 2009]* but only if the Multilinear Engine version 2 (ME2, *Paatero, 1999*) is run and processed separately from the EPA PMF Graphical User Interface (GUI). The procedure for modifying the PMF scripts to run the parametric model and the format of the output is described in Appendix A.

Section 3.1 describes the equations behind the expanded parametric model as implemented in the EPA PMF software and the approach we have used to integrate the data on wind speed, wind direction, season, and weekday/weekend effects into the model. Section 3.2 then describes the results of using this expanded parametric model to perform a factor analysis of the BIBE data for 2011 – 2014, comparing and contrasting the results of 6-, 7-, and 8-factor analyses using the expanded multilinear model with each other and the standard 6-factor PMF analysis from Section 2.

3.1 Equations for the Expanded Parametric Model

As noted above, in the procedure of *Paatero and Hopke [2002]*, the wind data is included as independent variables. Thus, the wind direction and wind speed data are not used directly, but are instead used to assign each day a wind direction and wind speed index.

For wind direction, this index (δ_i for day *i*) is based on the vector-averaged wind direction of the day. In *Paatero and Hopke [2002]* this index was the hourly average wind direction in degrees divided by 10 and rounded to the nearest integer. However, the guidance for the implementation of the expanded parametric model in the EPA PMF software recommends that that maximum wind speed index be no larger than 18. Thus, we used a wind direction index with values between 1 and 16, corresponding to the sixteen orthants (N = 1, NNE = 2, etc.) based on the data from KE38 (Alpine-Casparis Airport, to the north of the park, see Section 2.2).

For wind speed, the index ν_i is assigned based on a chosen classification of wind speeds. In *Paatero and Hopke [2002]*, the boundaries of this classification were 0, 1.5, 2.5, 3.5, 5.8, and >5.8 m/s. Thus there were 5 categories, with $\nu_{ih} = 2$ for hours with a wind speed between 1.5 and 2.5 m/s. We used a system of 4 wind speed categories with boundaries of 0, 0.25, 3.5, 7.5, >7.5 m/s. The first category was chosen to correspond with calm winds (<0.5 knots, or approximately 0.25 m/s), and the other three to divide the observed wind speed magnitudes into equal populations.

Each day is also assigned an index ω_i for the day of the week, with a value of 1 for weekdays and a value of 2 for weekends. The day is also assigned a seasonal index σ_i with a value between 1 and 6, with 1 for days in January or February, 2 for days in March or April, etc.

The goal is then to solve both the standard PMF equations

$$x_{ij} = \sum_{p=1}^{P} g_{ip} f_{jp} + e_{ij} \quad (1)$$

(where $p = 1,\ldots,P$ is the index of each source, j is the index for the measured chemical species, x_{ij} is the measured concentration of species j on day i, g_{ip} is the strength of source p on day i from the i by p matrix **G**, and f_{jp} is the concentration of species j in the emissions from source p from the j by p matrix **F**, and e_{ij} is the model residual for species j on day i) as well as the expanded model equations

$$0 = m_{ip} - g_{ip} + e'_{ip} \quad (2)$$

$$m_{ip} = \mathbf{D}(\delta_i, p)\mathbf{V}(v_i, p)\mathbf{W}(\omega_i, p)\mathbf{S}(\sigma_i, p) \quad (3)$$

where:

- **D** is a 16 by p matrix and $\mathbf{D}(\delta_i, p)$ is the single element of **D** that has the indices δ_i based on the wind direction and p based on the source
- **V** is a 4 by p matrix and $\mathbf{V}(v_i, p)$ is a single element of **V** that has the indices v_i based on the wind speed and p based on the source
- **W** is a 2 by p matrix and $\mathbf{W}(\omega_i, p)$ is a single element of **W** that has the indices ω_i based on the day of week and p based on the source
- **S** is a 6 by p matrix and $\mathbf{S}(\sigma_i, p)$ is a single element of **S** that has the indices σ_i based on the season and p based on the source.

Note again that in the above equations, the wind speed, wind direction, day of week, and seasonal data is *not* used directly in the equations. Rather, that data is used to derive indices, and those indices decide which elements of the unknown matrices **D**, **V**, **W**, and **S** are included in the above equation for a given day i. In addition, the weekday coefficients (row 1) of the 2 by p matrix **W** are fixed to 1, so that the second row gives the average strength of each factor on weekends and holidays, relative to weekdays.

The best-fit solution is defined as the unknown (but constrained to be positive) values of the unknown matrices **G**, **F, D, V, S,** and **W** that minimize the sum-of-squares value Q, defined by

$$Q = \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{e_{ij}}{\sigma_{ij}}\right)^2 + \sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{e'_{ip}}{\sigma'_{ip}}\right)^2 \quad (4)$$

where σ_{ij} is the uncertainty in the measurement of species j on day i. The error estimates σ'_{ip} for the expanded equations must be much larger than σ_{ij} as the expanded model equations are

expected to give a poorer fit to the data. Further details on the mathematics of the solution as implemented in the EPA PMF software are given in *Paatero [2009]*.

3.2 Results

3.2.1 6-factor Expanded Parametric Model

Factor Identification

We first used the expanded parametric model to fit six factors, as this allows us to see how the addition of these meteorological variables affects the final solution relative to the 6-factor baseline PMF analysis presented in Section 2. The species profiles for the six factors are presented in Figure 14 to Figure 19. We have made the following identifications based on these profiles:

- Factor 1: Given the relatively high amounts of sulfate and nitrate, along with significant amounts of sodium, we think this factor is a mixture of sea salt and secondary anthropogenic pollution, which we refer to as "Pollution/Sea Salt" hereafter.
- Factor 2: This factor has high amounts of Si, Fe, and Ca, and so appears to be natural mineral dust, or "Dust (Fe, Si, Ca)".
- Factor 3: Given the relatively high amounts of sulfate and nitrate, along with high amounts of lead (Pb) and arsenic (As), we identify this factor as a mixture of secondary sulfate pollution with lead and arsenic, or "Sulfate with Pb, As".
- Factor 4: The high levels of organic carbon (OC) and elemental carbon (EC), combined with the relatively high potassium (K) level, suggests that this factor is smoke from biomass burning, or "Smoke".
- Factor 5: This factor mainly includes only one type of OC (OC1) and makes a small contribution to the overall mass balance. Thus we label this factor as "Trace OC".
- Factor 6: This factor has high sodium (Na), chlorine (Cl), and magnesium (Mg) content, suggesting that it is from sea salt, which we call "High Mg Sea Salt" to distinguish it from the "Pollution/Sea Salt" factor.

We note that, while the standard 6-factor PMF separated the sulfate and Pb/As contributions into different factors, the extended parametric model groups them into the same factor here. The extended parametric model also distributes the nitrate contributions more evenly between the factors, making it difficult to identify a factor where nitrate is separate from sulfate.

Dependence on Wind, Seasonality, and Day of Week

The parametric factor matrices fit by the extended parametric model (i.e., **D, V, S,** and **W** in Equation 3) can also be useful in determining if our identifications of the above factors are consistent with their dependence on wind direction, wind speed, season, and weekend/weekday dependence. Figure 20 shows a radar plot of the dependence of each factor on wind direction. The "Dust (Fe, Si, Ca)" factor is enhanced when the winds are from the northwest, consistent with this being wind-blown mineral dust from desert regions. We can also see that the pollution factors, "Pollution/Sea Salt" and "Sulfate with Pb, As" are enhanced when the winds are from the south to east, consistent with this being anthropogenic pollution from Mexico or Southeast Texas. "Smoke" is somewhat enhanced when the winds are from the south, suggesting that biomass burning in Mexico is contributing significantly to this factor. However, "High Mg Sea Salt" does not have a strong dependence on wind direction, which may reflect equal long-distance transport from the Pacific in the west and the Gulf of Mexico in the East and South.

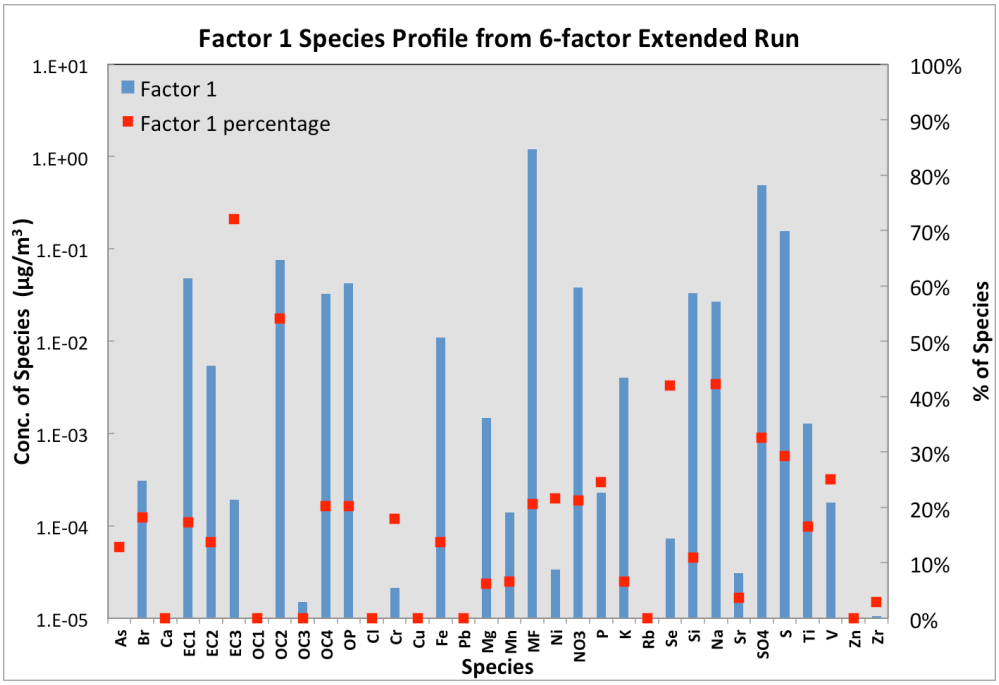

Figure 14. Species profile for factor 1, identified as a mixture of secondary pollution and sea salt ("Pollution/Sea Salt"), from the 6-factor extended parametric model fit. Blue bars show the normalized concentration of the species in the factor, while the red dots show the percentage of that species in the factor.

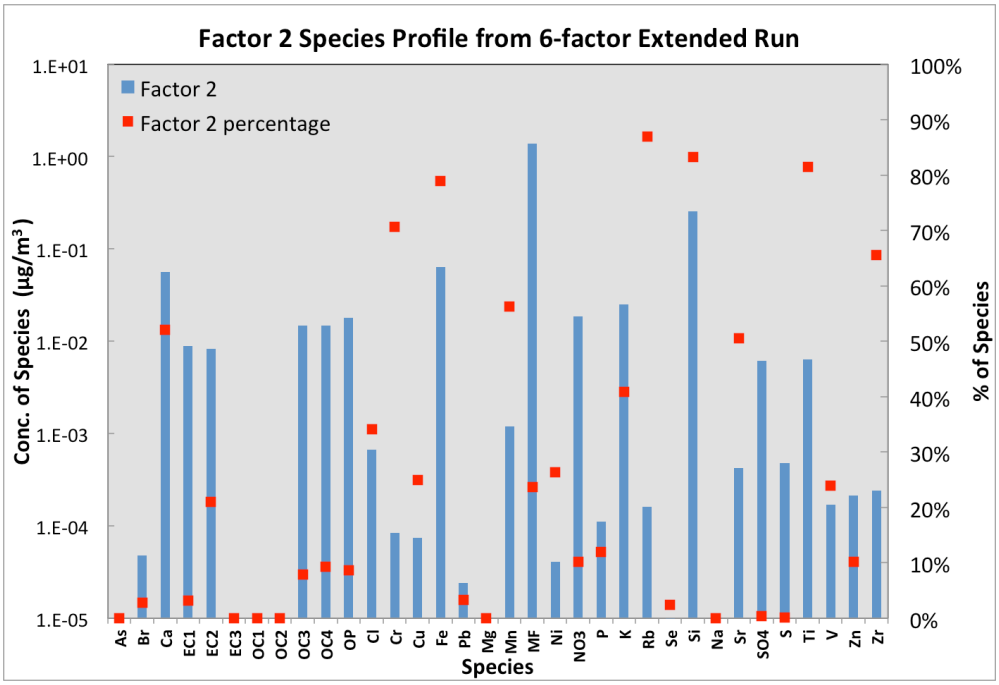

Figure 15. Species profile for factor 2, identified as natural mineral dust ("Dust (Fe, Si, Ca)"), from the 6-factor extended parametric model fit.

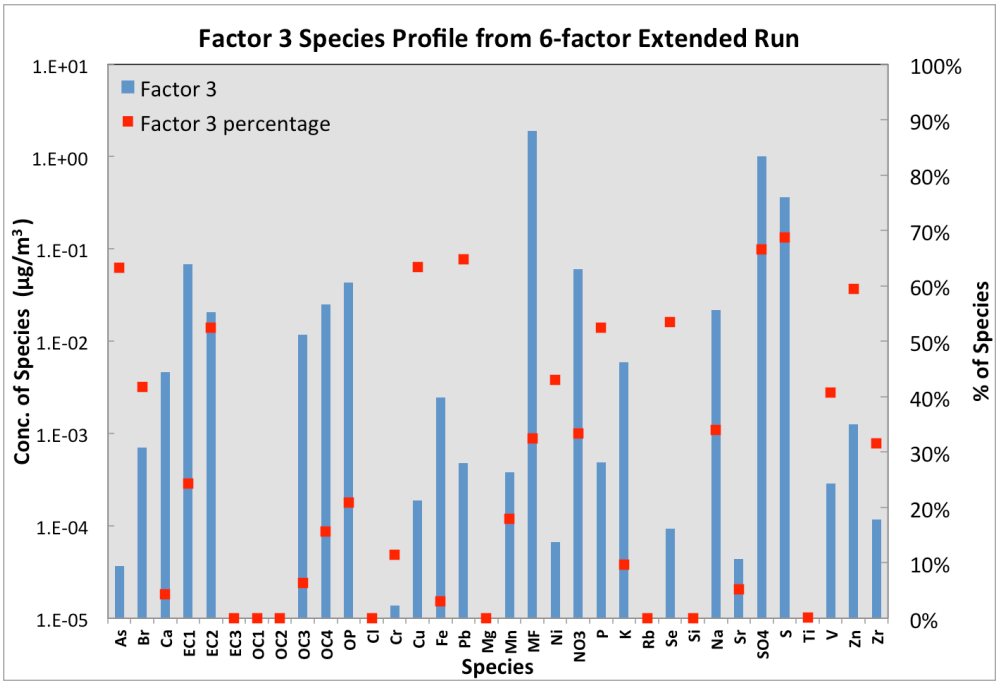

Figure 16. Species profile for factor 3, identified as a mixture of secondary sulfate pollution with lead and arsenic ("Sulfate with Pb, As"), from the 6-factor extended parametric model fit.

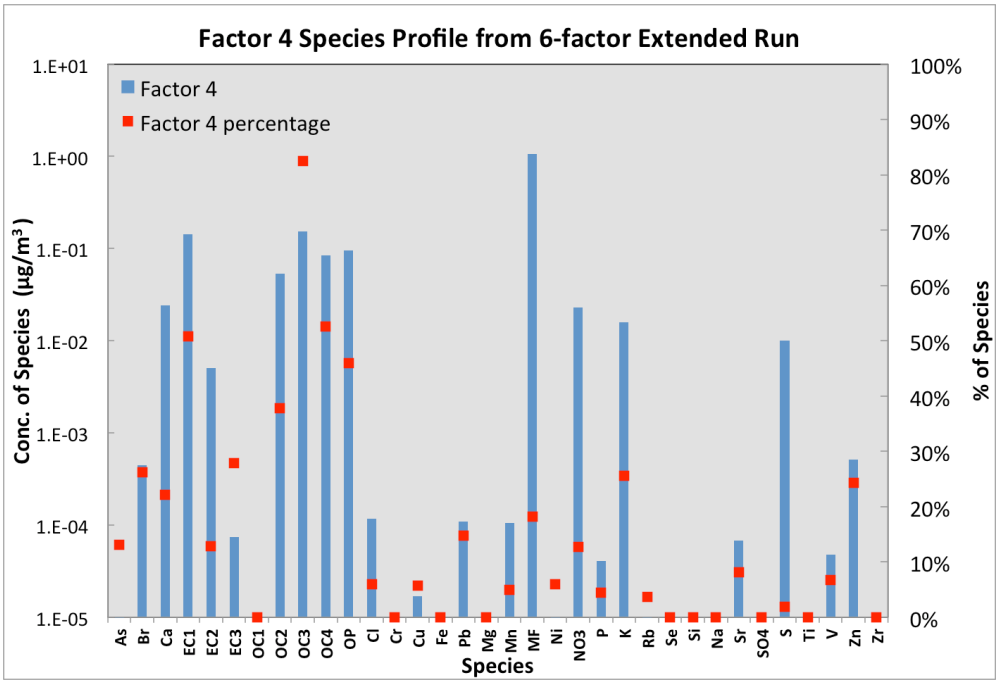

Figure 17. Species profile for factor 4, identified as smoke from biomass burning ("Smoke"), from the 6-factor extended parametric model fit.

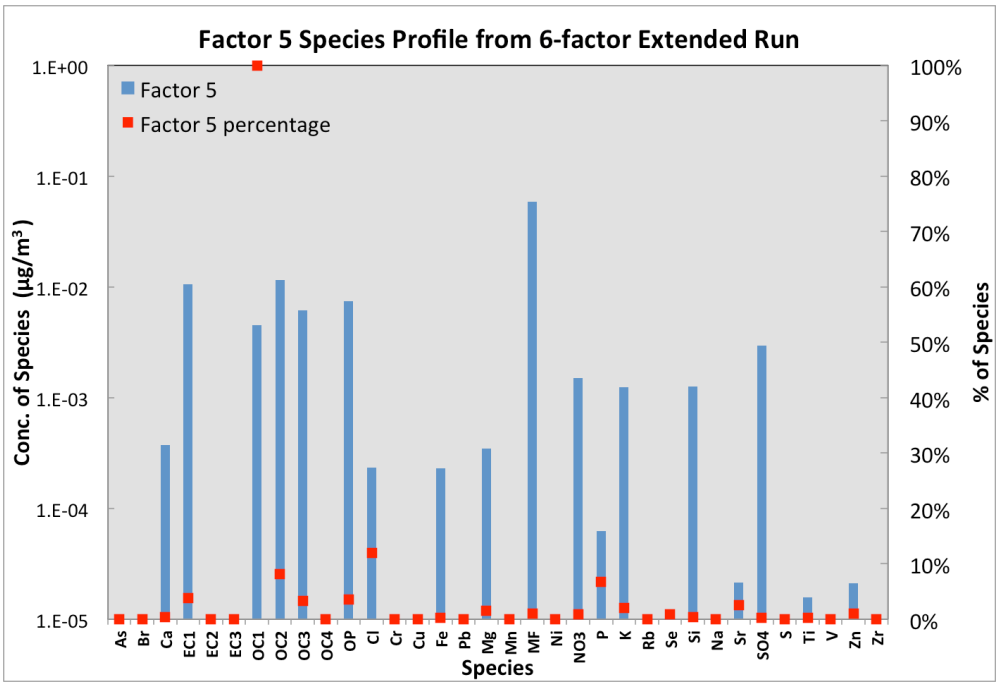

Figure 18. Species profile for factor 5, identified as trace organic carbon ("Trace OC"), from the 6-factor extended parametric model fit.

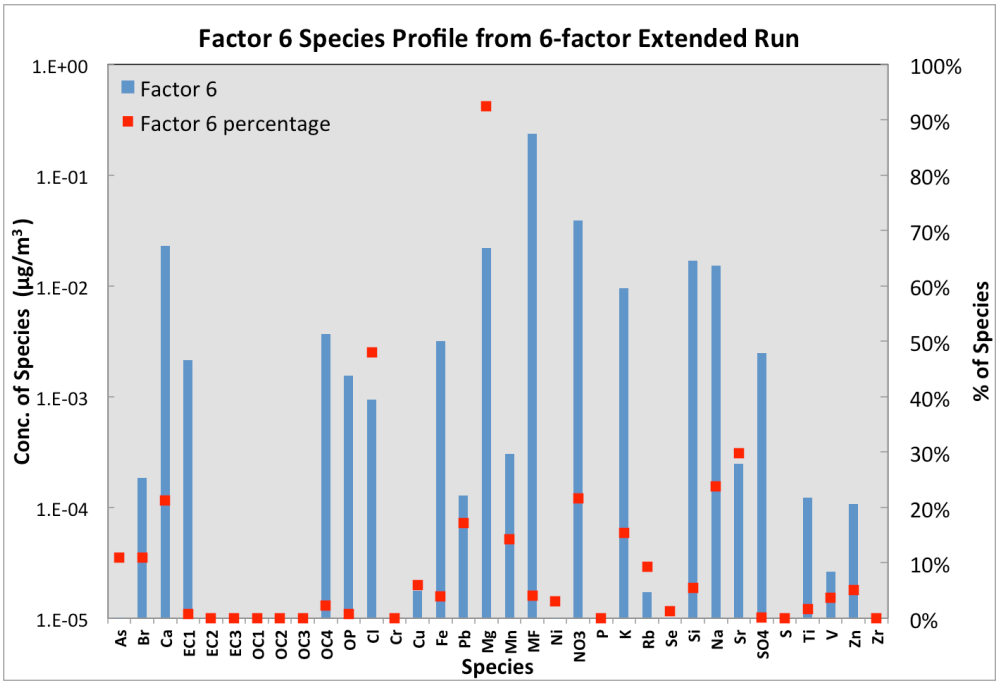

Figure 19. Species profile for factor 6, identified sea salt with a high magnesium content ("High Mg Sea Salt"), from the 6-factor extended parametric model fit.

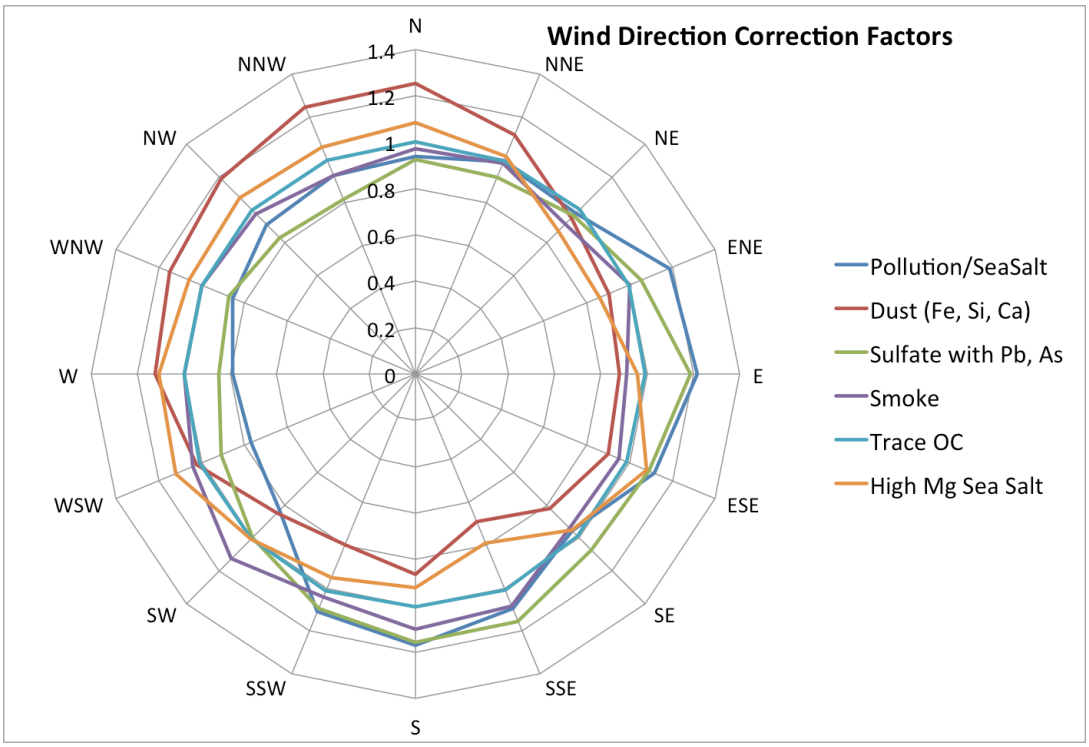

Figure 20. Wind direction correction factors for the 6-factor extended parametric model fit. A value above 1 means that the factor is enhanced when the wind is blowing from the given direction.

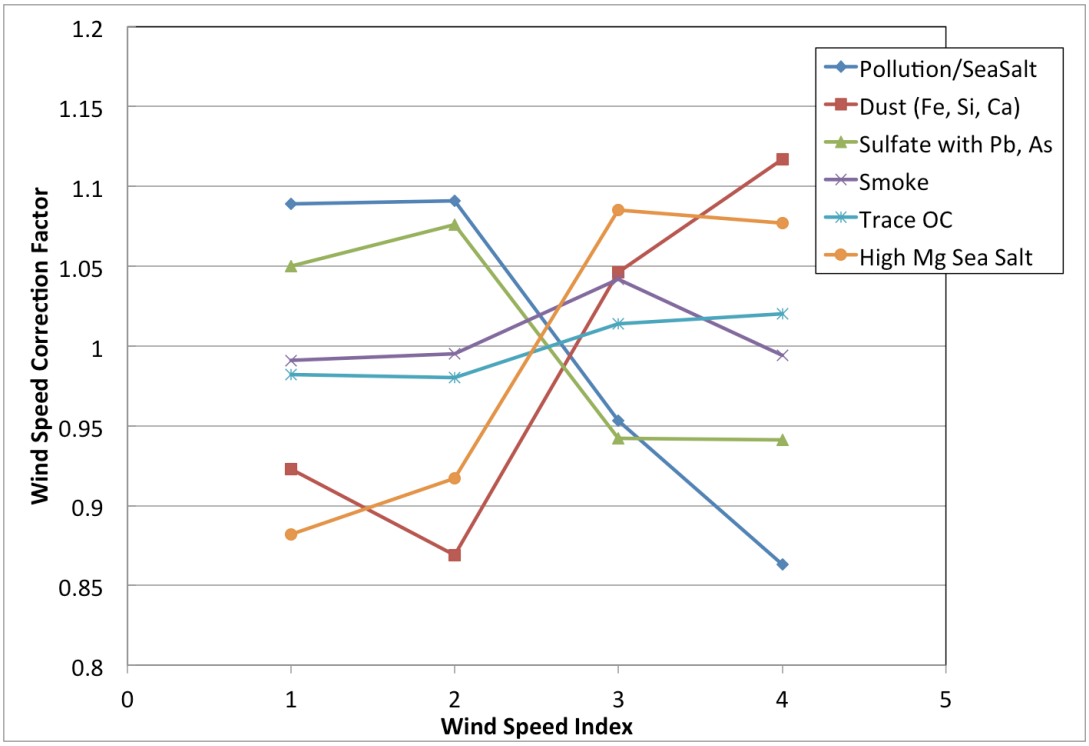

Figure 21. Wind speed correction factors for the 6-factor extended parametric model fit. A value above 1 means that the factor is enhanced when the wind is blowing at the give speed.

Figure 21 shows the dependence of the factors on wind speed. We can see that "Dust (Fe, Si, Ca)" and "High Mg Sea Salt" increase with wind speed, as expected since the emissions from these sources should be highest at high wind speeds. The influence of the two pollution factors decreases with wind speed, which could be due to greater dilution of the anthropogenic emissions leading to these pollutants at high wind speeds, but it does imply that sea salt is a minor component of the "Pollution/Sea Salt" factor. "Smoke" and "Trace OC" show little dependence with wind speed.

Figure 22 shows the seasonality of the different factors. "Dust (Fe, Si, Ca)" and "High Mg Sea Salt" have a strong dependence on season, with strongly enhanced values (correction factor ~1.4) in March and April, and weak values (correction factor ~0.5) in November-December. This seasonality does not appear to be linked to mean synoptic wind, but could potentially be linked to individual synoptic episodes. The "Pollution/Sea Salt" factor is also strongly dependent on season, but with a peak in May-June and a minimum in January-February. The "Sulfate with Pb, As" factor has a more moderate seasonal cycle, but still peaks in the summer, consistent with secondary anthropogenic pollution. "Smoke" peaks in May-June, which is somewhat consistent with the fire season in the Yucatan, which peaks in April-May.

Figure 23 shows the weekend correction factors for the factors. The only factor with a strong dependence on the day of the week appears to be "Dust (Fe, Si, Ca)", which is weaker on weekends, but it is difficult to understand why dust would depend on the day of the week.

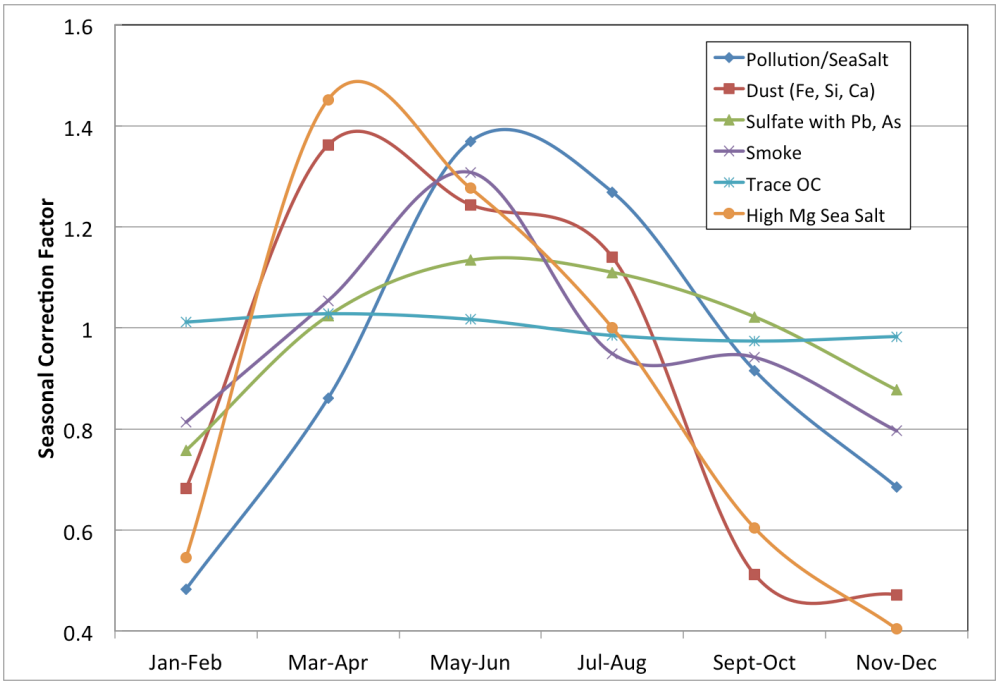

Figure 22. Seasonal correction factors for the 6-factor extended parametric model fit. A value above 1 means that the factor is enhanced during that two-month period.

Figure 23 shows the weekend corrections factors for each model.

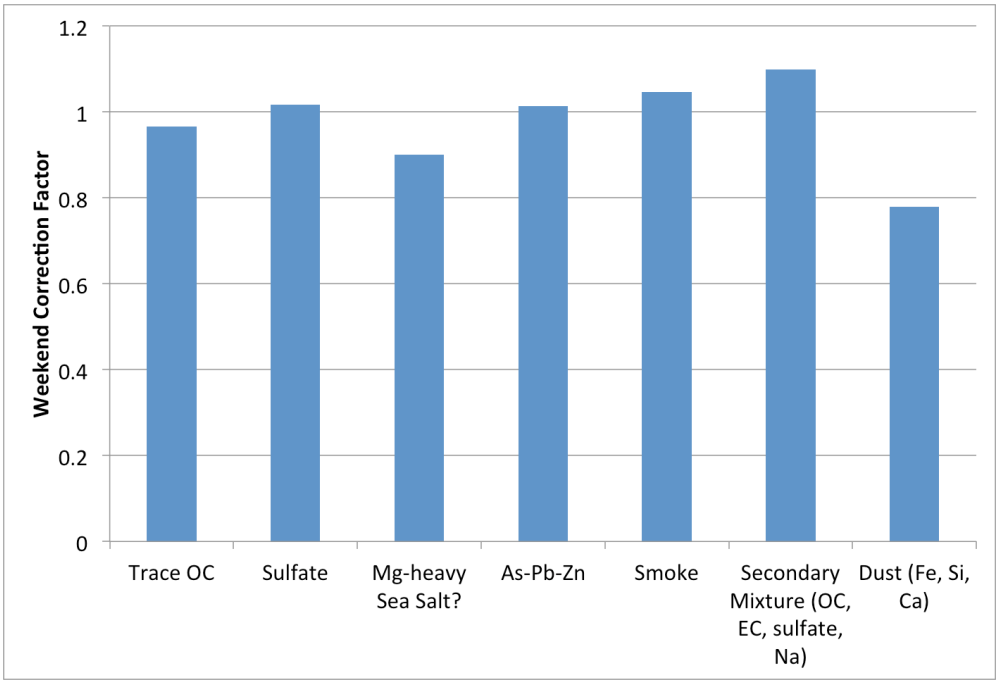

Figure 23. Weekend correction factors for the 6-factor extended parametric model fit. A value above 1 means that the factor is enhanced during that two-month period.

Contributions of Each Factor to $PM_{2.5}$ at BIBE

Figure 24 shows the relative contributions of the factors from the 6-factor extended parametric model fit averaged over all days at BIBE with IMPROVE aerosol measurements between 2011-2014. The estimated contributions from dust (24 %) and smoke (18 %) are consistent with the estimated contributions from the original 6-factor PMF analysis in Section 2 (24 % and 17 %, respectively). The "Sulfate with Pb, As" factor here is slightly less strong than the "Sulfate" factor from the original PMF (32 % versus 43 %, respectively), but this appears to be because the "Pollution/Sea Salt" mixture is accounting for the balance of the anthropogenic pollution contribution.

The relative contributions of the factors to the days above the 80th percentile for $PM_{2.5}$ at BIBE (the "Top 20 %") are shown in Figure 25. The contributions from dust and smoke increase to 30 % and 20 %, respectively, while the contributions from the pollution factors decrease. Interestingly, smoke also makes a larger than average contribution to the days below the 20th percentile for $PM_{2.5}$ (25 %, Figure 26), while dust has a smaller than average contribution to these days (20%). In all cases, the two pollution factors contribute about 44-53 % to the total $PM_{2.5}$.

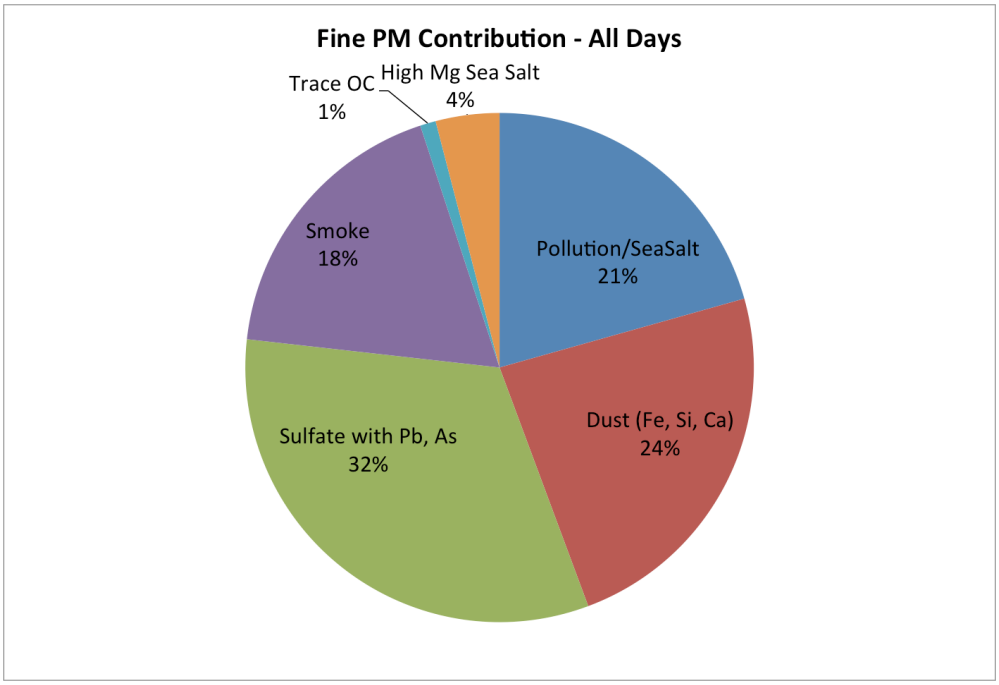

Figure 24. Relative contributions of the factors from the 6-factor extended parametric model fit averaged over all days at BIBE with IMPROVE aerosol measurements between 2011-2014.

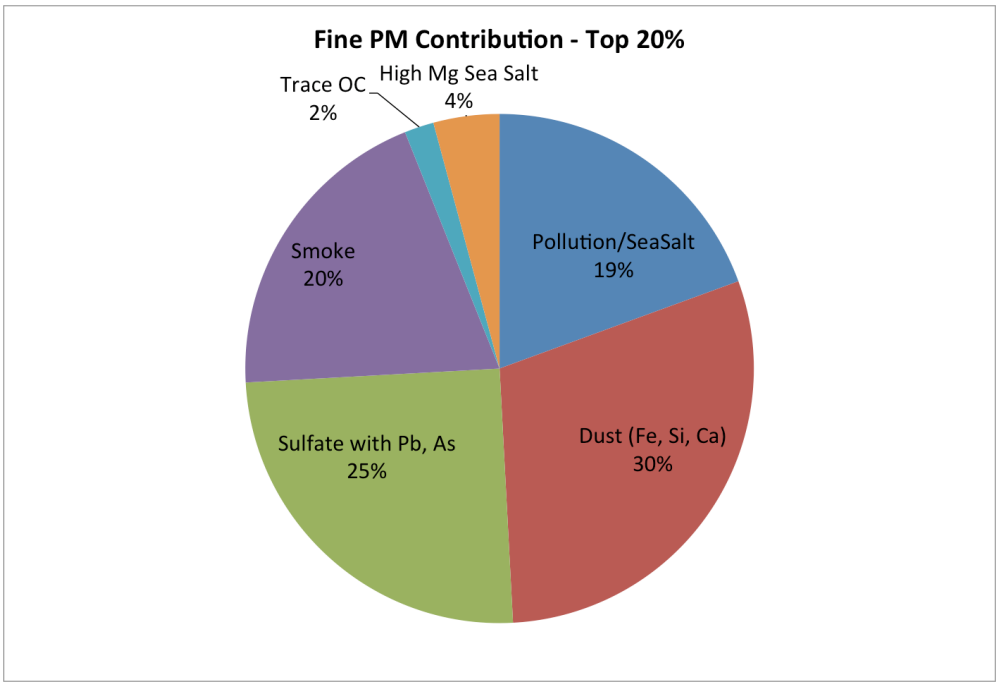

Figure 25. As in Figure 24, but only for the days above the 80th percentile for $PM_{2.5}$ at BIBE (the "Top 20%").

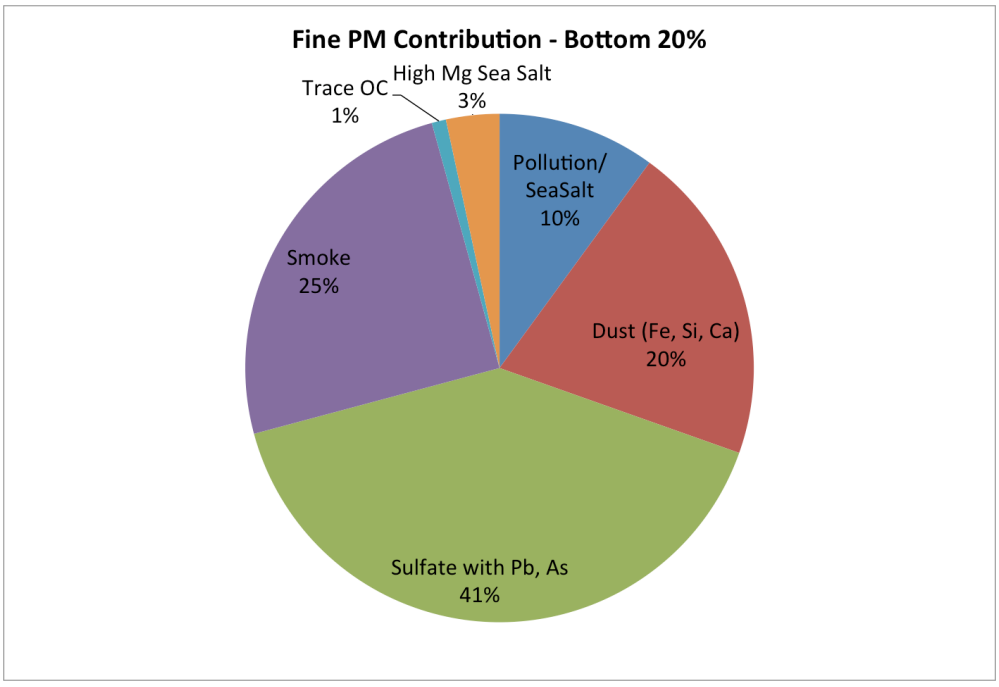

Figure 26. As in Figure 24, but only for the days below the 20th percentile for $PM_{2.5}$ at BIBE (the "Bottom 20%").

Figure 27 to Figure 30 shows the time series of the different factor's absolute contributions to $PM_{2.5}$ at BIBE for each of the four years analyzed. We can see strong episodic contributions of the "Smoke" factor in 2011, which was a severe drought year in Texas, as well as episodic high dust contributions due to dust storms. Contributions from the other factors appear show less day-to-day and inter-annual variability.

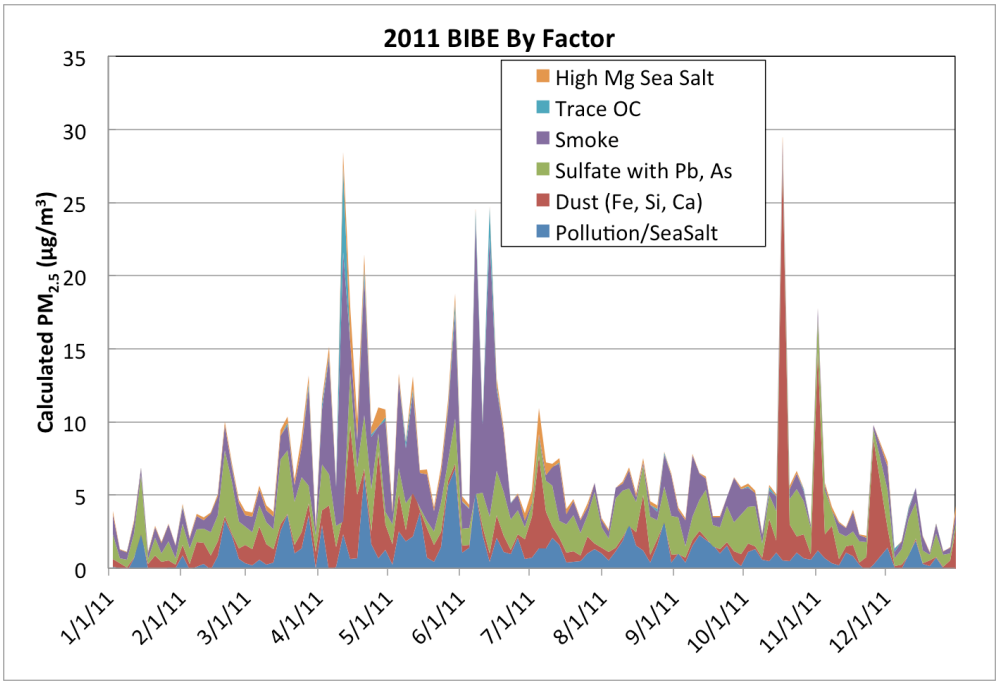

Figure 27. Time series of the absolute ($\mu g/m^3$) contributions of the factors from the 6-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2011.

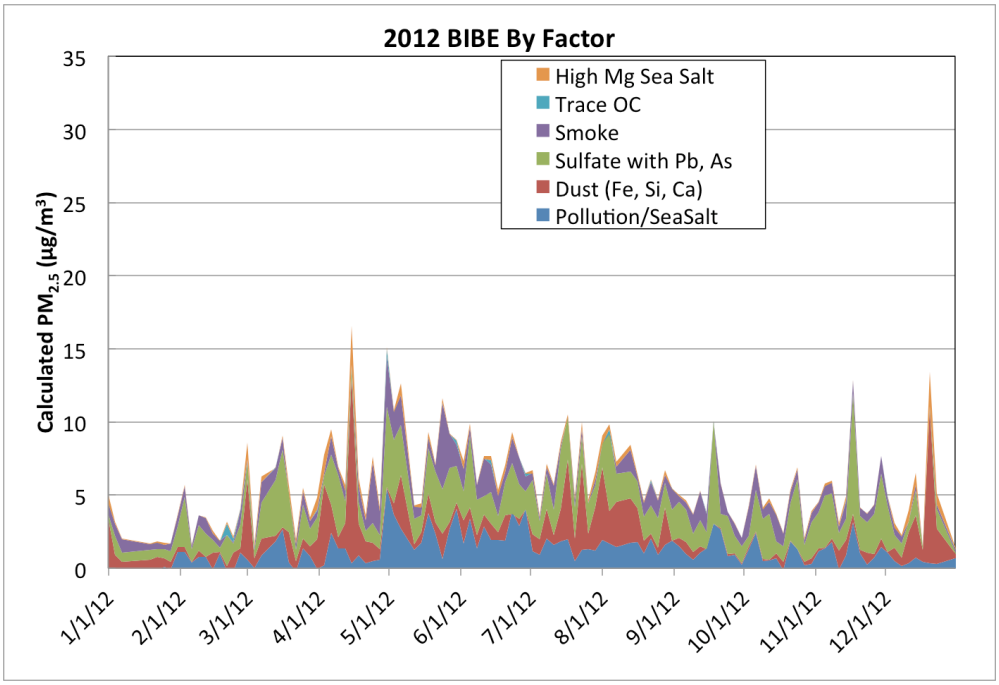

Figure 28. Time series of the absolute (µg/m^3) contributions of the factors from the 6-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2012.

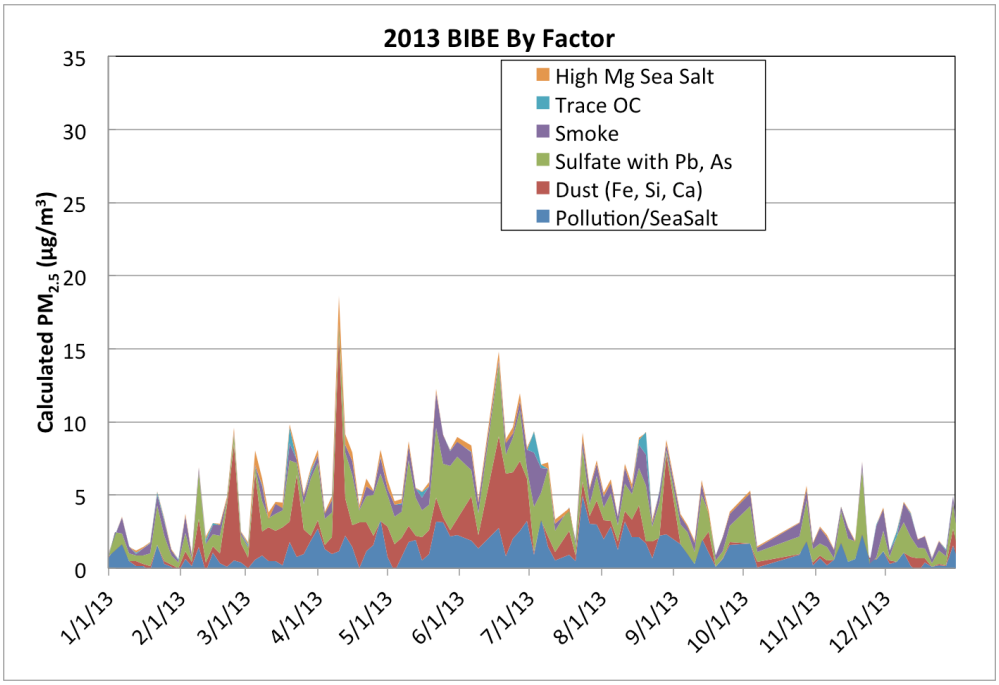

Figure 29. Time series of the absolute (µg/m^3) contributions of the factors from the 6-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2013.

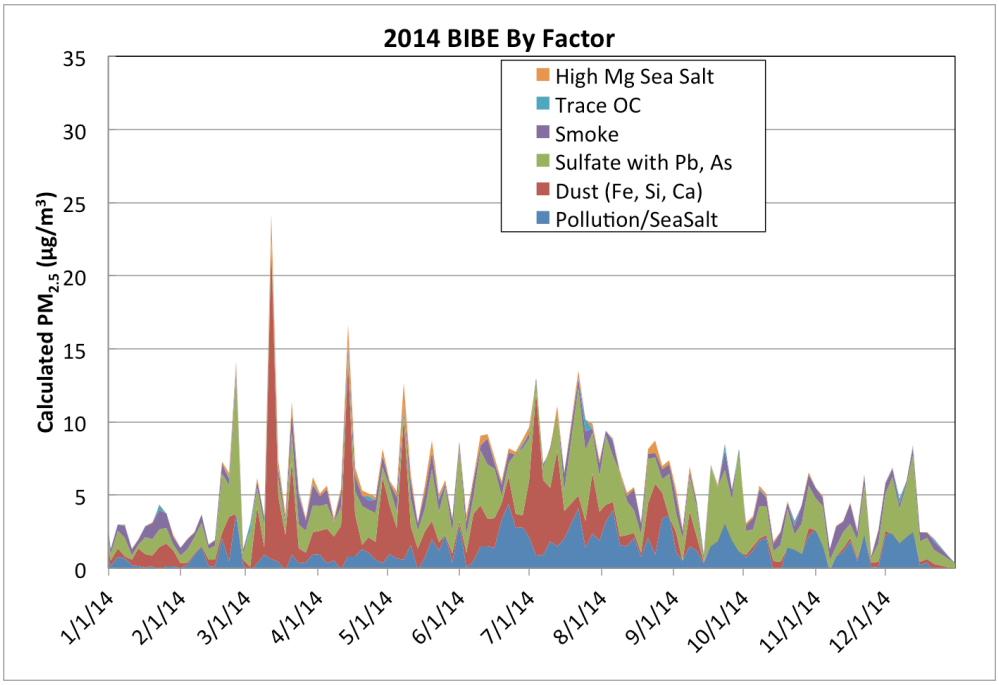

Figure 30. Time series of the absolute (µg/m^3) contributions of the factors from the 6-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2014.

3.2.2 7-factor Expanded Parametric Model

Factor Identification

Part of the motivation of adding meteorological data to a PMF analysis is to see if that allows for factors to be better separated than the basic PMF. Thus we ran the extended parametric model again, this time attempting to fit 7 factors.

The species profiles for the seven factors are presented in Figures 31 to 37. We have made the following identifications based on these profiles:

- Factor 1: This factor mainly includes only one type of OC (OC1) and makes a small contribution to the overall mass balance. Thus we label this factor as "Trace OC", like Factor 5 from the 6-factor expanded parametric model fit.
- Factor 2: This factor is predominantly sulfate from anthropogenic pollution ("Sulfate").
- Factor 3: This factor has high sodium (Na), chlorine (Cl), and magnesium (Mg) content, suggesting that it is from sea salt, which we call "High Mg Sea Salt" to distinguish it form the "Pollution/Sea Salt" factor, like Factor 2 from the 6-factor expanded parametric model fit.
- Factor 4: This factor has high relative contributions of heavy metals like arsenic (As), copper (Cu), lead (Pb), and zinc (Zn) and so is labeled "As-Cu-Pb-Zn" until a more precise identification can be made.
- Factor 5: The high levels of organic carbon (OC) and elemental carbon (EC), combined with the relatively high potassium (K) level, suggests that this factor is smoke from biomass burning, or "Smoke", like Factor 2 from the 6-factor expanded parametric model fit.

- Factor 6: This factor has sulfate, nitrate, and sodium, as well as OC and EC, and thus is labeled "Mixed Pollution".
- Factor 7: This factor has high amounts of Si, Fe, and Ca, and so appears to be natural mineral dust, or "Dust (Fe, Si, Ca)", like Factor 2 from the 6-factor expanded parametric model fit.

We can see the main effect of adding an additional factor to the model was the separation of "Sulfate" from the heavy-metal "As-Cu-Pb-Zn" factor. Furthermore, the "Mixed Pollution" factor is from this fit is very well correlated with the "Sea Salt/Nitrate" factor from the 6-factor parametric fit (R^2 of 0.93), suggesting these represent the same source but the contribution from this source is estimated to be about 30 % smaller in the 7-factor parametric fit (Figure 25).

Our results are also generally consistent with the baseline 6-factor PMF without the added meteorological data (Section 2). Table 2 shows the correlations and ordinary-least-squares regression slopes between the concentrations of the factors from the 6-factor baseline PMF (x) and the 7-factor extended parametric fit (y). "All Dust" is the sum of the mineral and heavy metal dust factors. The smoke, sulfate, mineral dust, mixed pollution, and all dust factors are very well correlated between the two analyses ($R^2 > 0.8$), while the heavy metal dust and trace nitrate and trace OC factors show little correlation. However, these factors are minor contributors to the overall $PM_{2.5}$ at BIBE, and so we conclude that our results are generally robust across these two PMF approaches.

Table 2. Linear correlation analysis of the 6-factor baseline PMF factors (x) and the 7-factor extended parametric fit (y).

Baseline Factor	Parametric Factor	R^2	Slope
Dust (Fe, Si)	Dust (Fe, Si, Ca)	0.97	0.88
Dust (As, Pb)	As-Cu-Pb-Zn	0.39	2.43
All Dust	All Dust	0.94	0.89
Smoke	Smoke	0.86	0.81
Sulfate	Sulfate	0.83	1.02
Sea Salt/Nitrate	Mixed Pollution	0.93	0.71
Sea Salt/Nitrate	High Mg Sea Salt	0.18	0.15
Nitrate	Trace OC	0.06	0.69

We also attempted an 8-factor parametric fit (not shown). However, this appeared to mix the factors together in ways that made them difficult to identify. We thus think that the 7-factor fit gives the best results for the for the expanded parametric model.

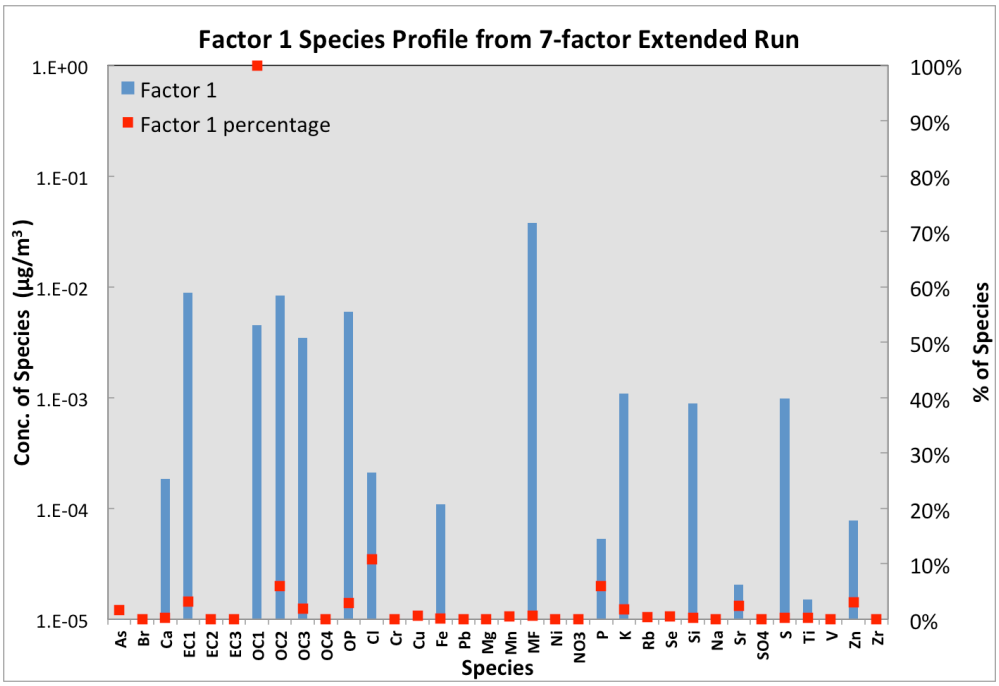

Figure 31. Species profile for factor 1, identified as trace organic carbon ("Trace OC"), from the 7-factor extended parametric model fit. Blue bars show the normalized concentration of the species in the factor, while the red dots show the percentage of that species in the factor.

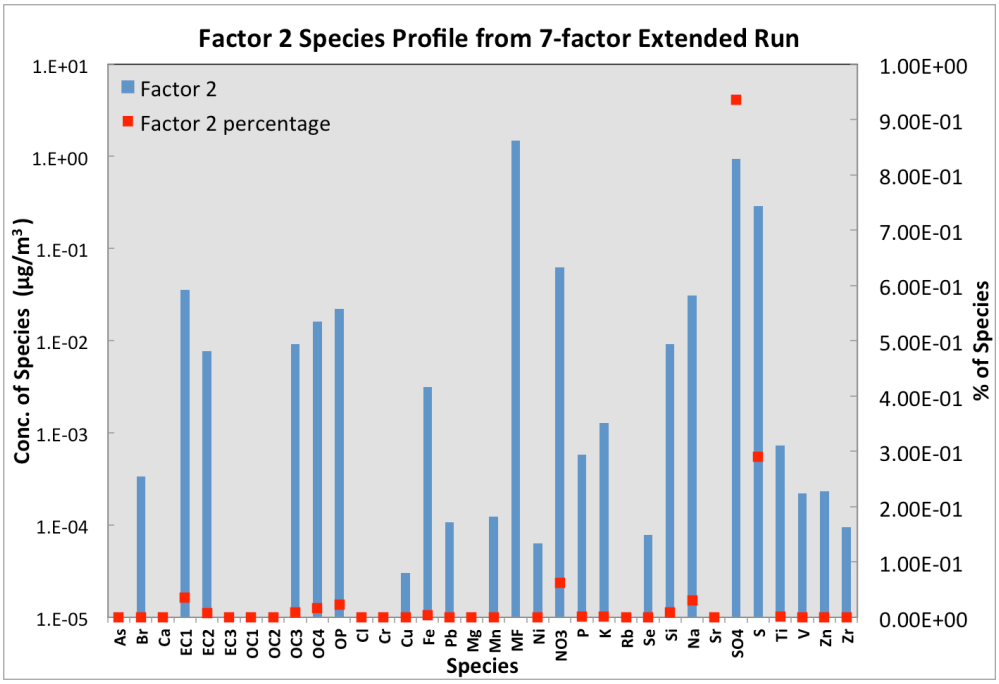

Figure 32. Species profile for factor 2, identified as secondary sulfate pollution ("Sulfate"), from the 7-factor extended parametric model fit.

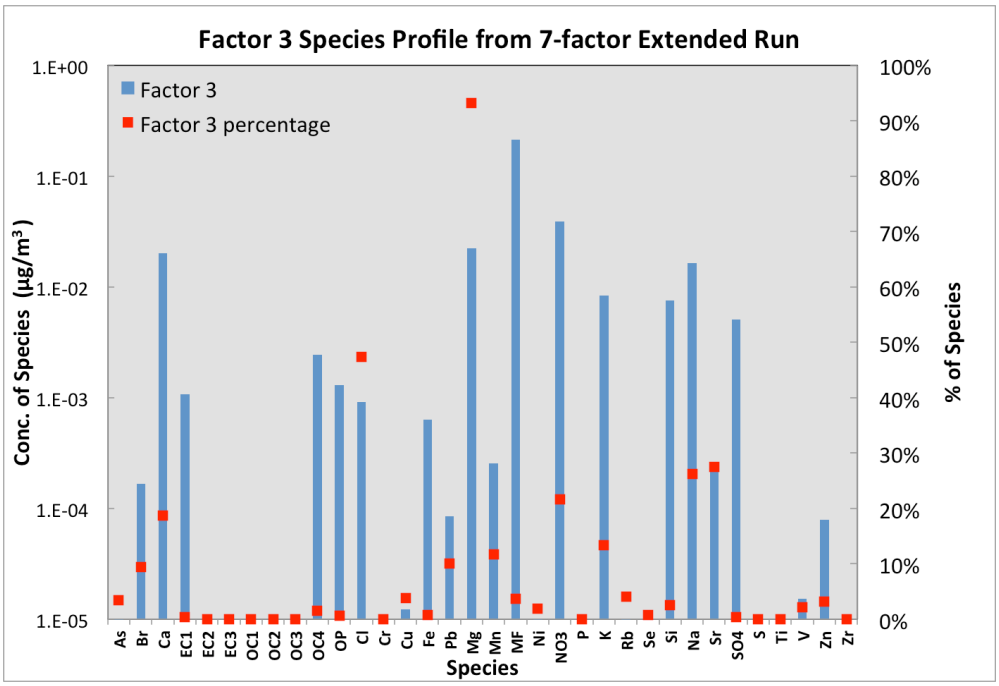

Figure 33. Species profile for factor 3, identified as sea salt with a high magnesium content ("High Mg Sea Salt"), from the 7-factor extended parametric model fit.

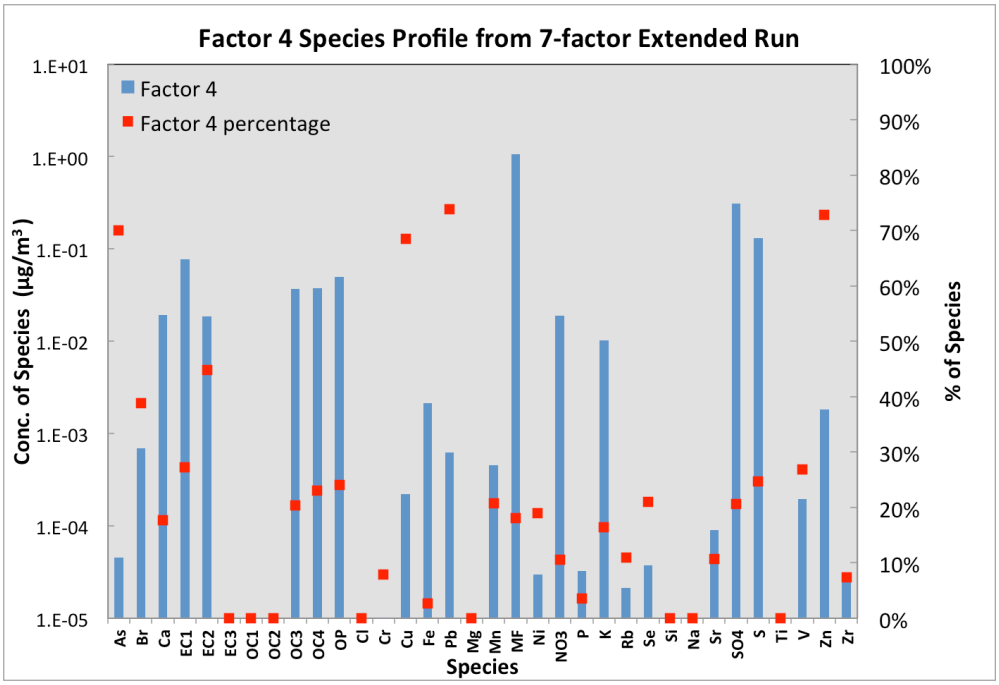

Figure 34. Species profile for factor 4, identified as heavy metal dust ("As-Cu-Pb-Zn"), from the 7-factor extended parametric model fit.

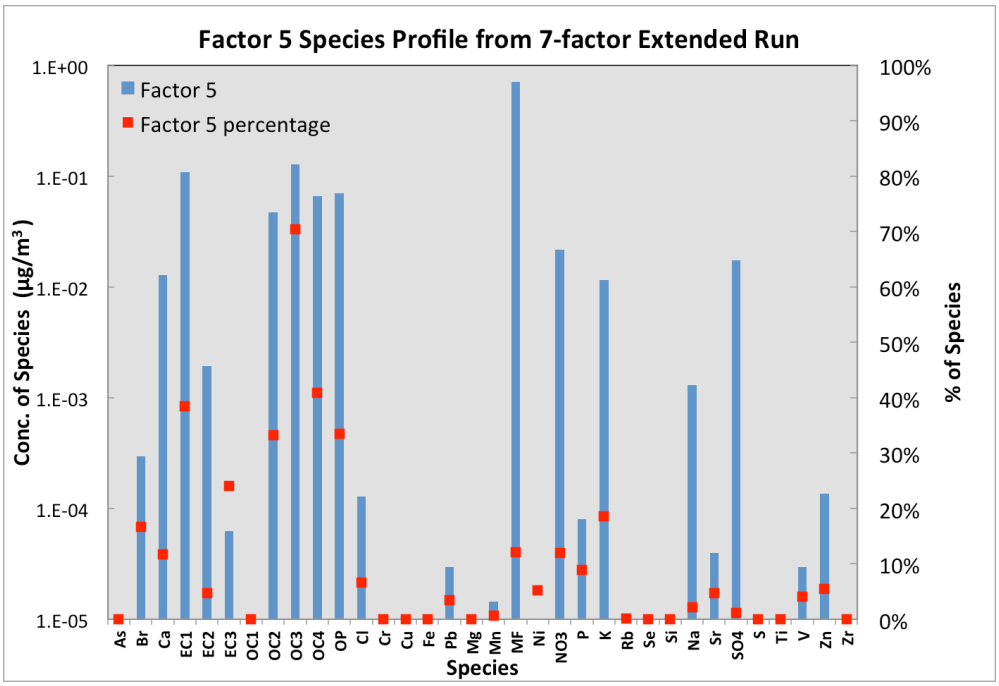

Figure 35. Species profile for factor 5, identified as smoke from biomass burning ("Smoke"), from the 7-factor extended parametric model fit.

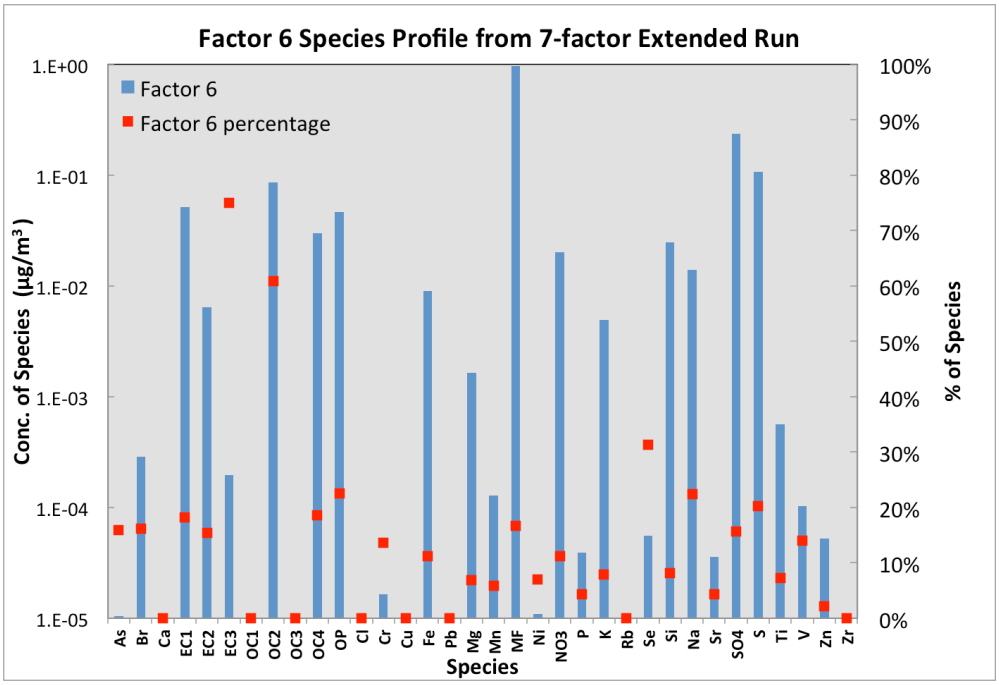

Figure 36. Species profile for factor 6, identified as a mixture of secondary pollution ("Mixed Pollution") from the 7-factor extended parametric model fit.

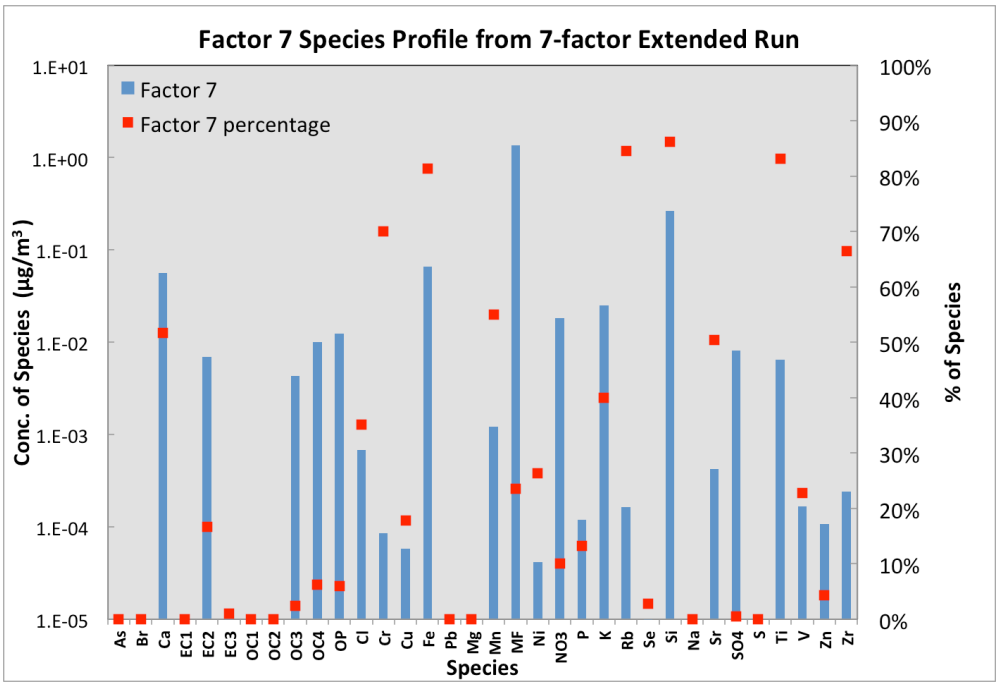

Figure 37. Species profile for factor 7, identified as natural mineral dust ("Dust (Fe, Si, Ca)"), from the 7-factor extended parametric model fit.

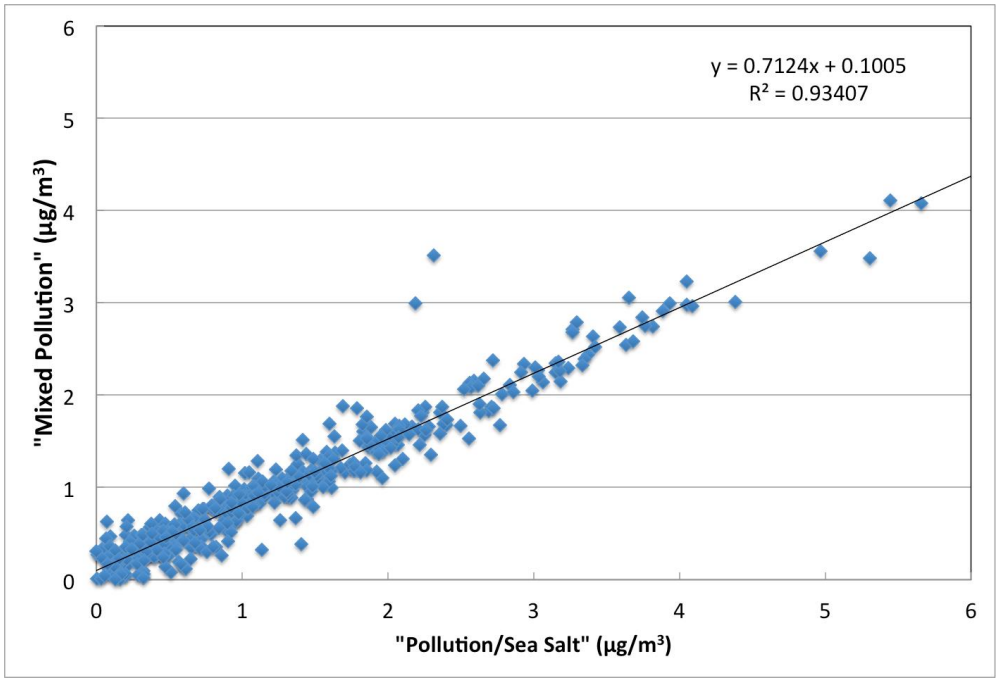

Figure 38. Correlation between the "Pollution/Sea Salt" factor from the 6-factor extended parametric model fit with the "Mixed Pollution" factor from the 7-factor fit.

Dependence on Wind, Seasonality, and Day of Week

Figure 39 shows a radar plot of the dependence of each factor on wind direction. Again, the "Dust (Fe, Si, Ca)" factor is enhanced when the winds are from the northwest, consistent with this being wind-blown mineral dust from desert regions. The enhancement of the pollution factors "Sulfate" and "Mixed Pollution" when the winds are from the South (Mexico) and East (Texas) is even clearer in this fit. We can also see that the pollution factors, "Pollution/Sea Salt" and "Sulfate with Pb, As" are enhanced, consistent with this being anthropogenic pollution from Mexico or Southeast Texas. However, "Smoke" and the other factors don't show significant dependence on direction in this fit.

Figure 40 shows the dependence of the factors on wind speed. As before, "Dust (Fe, Si, Ca)" and "High Mg Sea Salt" increase with wind speed, and the two pollution factors ("sulfate" and "Mixed Pollution") decrease with wind speed. "Smoke" and "Trace OC" show little dependence with wind speed. The other factors all show a moderate increase with wind speed.

Figure 41 shows the seasonality of the different factors. As before, "Dust (Fe, Si, Ca)" and "High Mg Sea Salt" have a strong dependence on season, with strongly enhanced values (correction factor ~1.4) in March and April, and weak values (correction factor ~0.5) in November-December. The "Sulfate" and "Mixed Pollution" factors are strongly dependent on season, peaking in the summer consistent with secondary anthropogenic pollution. "Smoke" again peaks in May-June, which is somewhat consistent with the fire season in the Yucatan, which peaks in April-May. The "As-Cu-Pb-Zn" factor shows a moderate seasonality, peaking in March-April and September-October. As before, the only factor with a strong dependence on the day of the week appears to be "Dust (Fe, Si, Ca)", which is weaker on weekends (Figure 42).

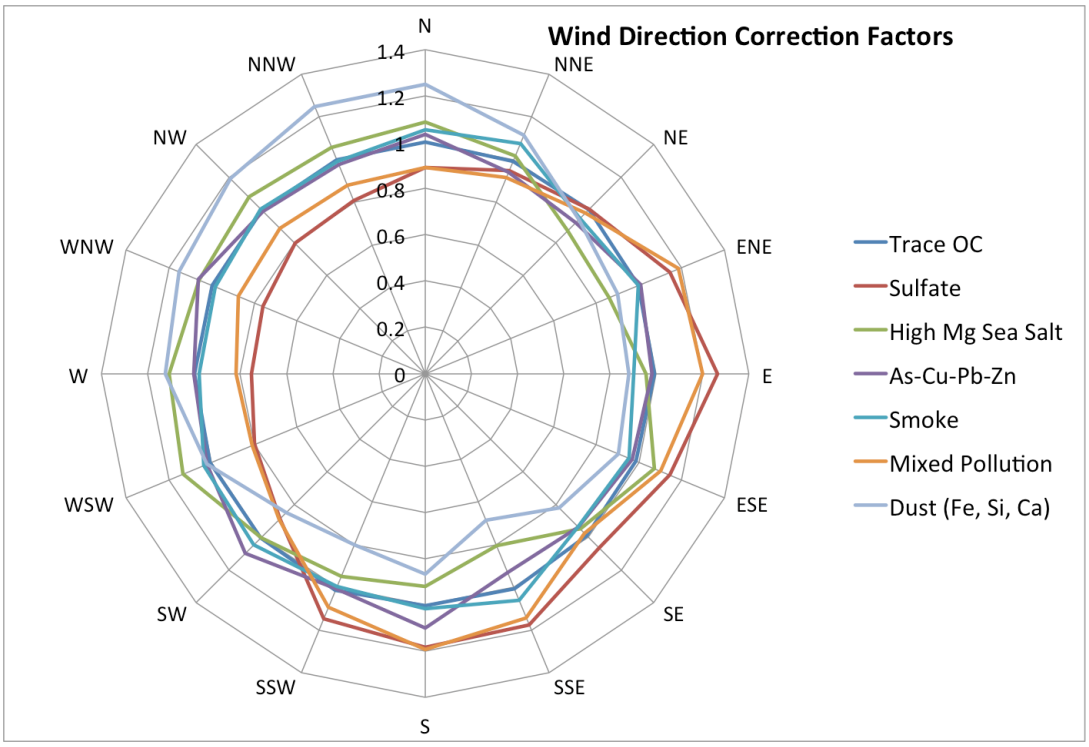

Figure 39. Wind direction correction factors for the 7-factor extended parametric model fit. A value above 1 means that the factor is enhanced when the wind is blowing from the given direction.

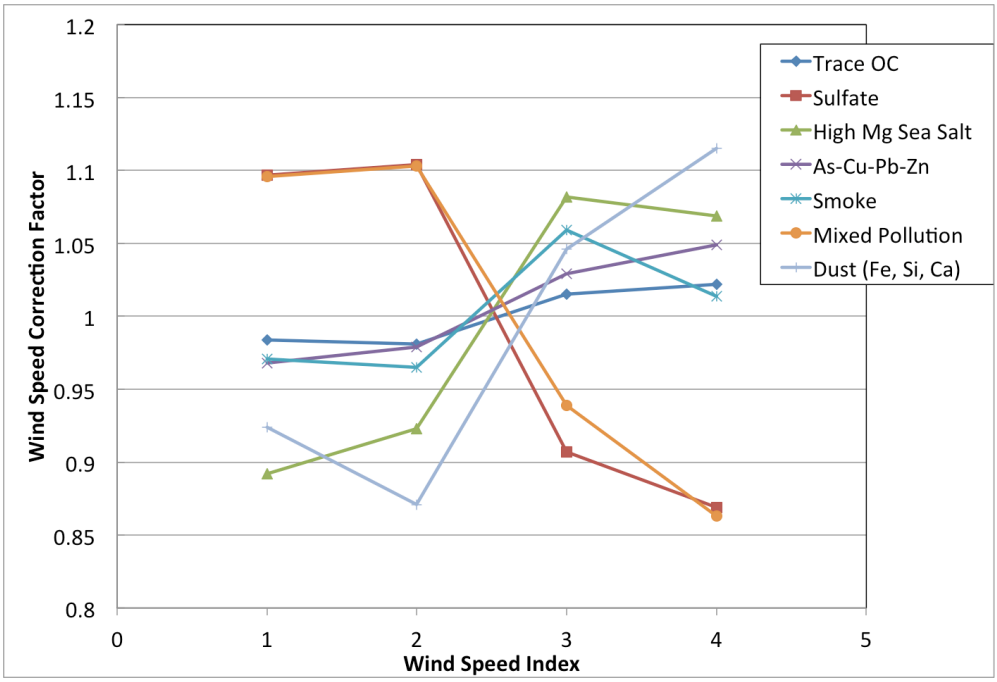

Figure 40. Wind speed correction factors for the 7-factor extended parametric model fit. A value above 1 means that the factor is enhanced when the wind is blowing at the give speed.

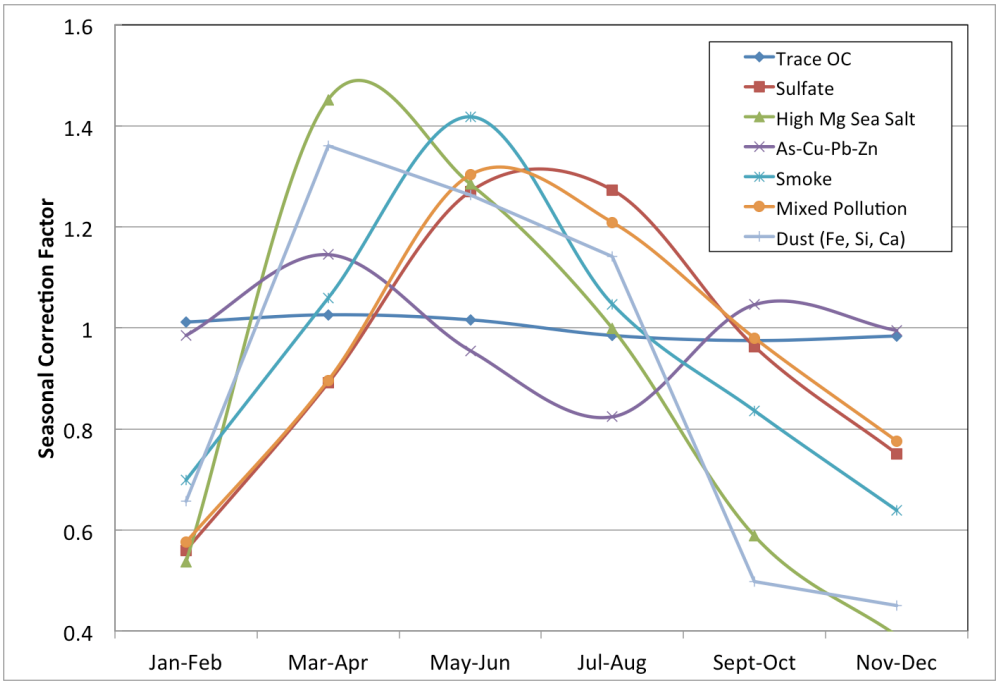

Figure 41. Seasonal correction factors for the 7-factor extended parametric model fit. A value above 1 means that the factor is enhanced during that two-month period.

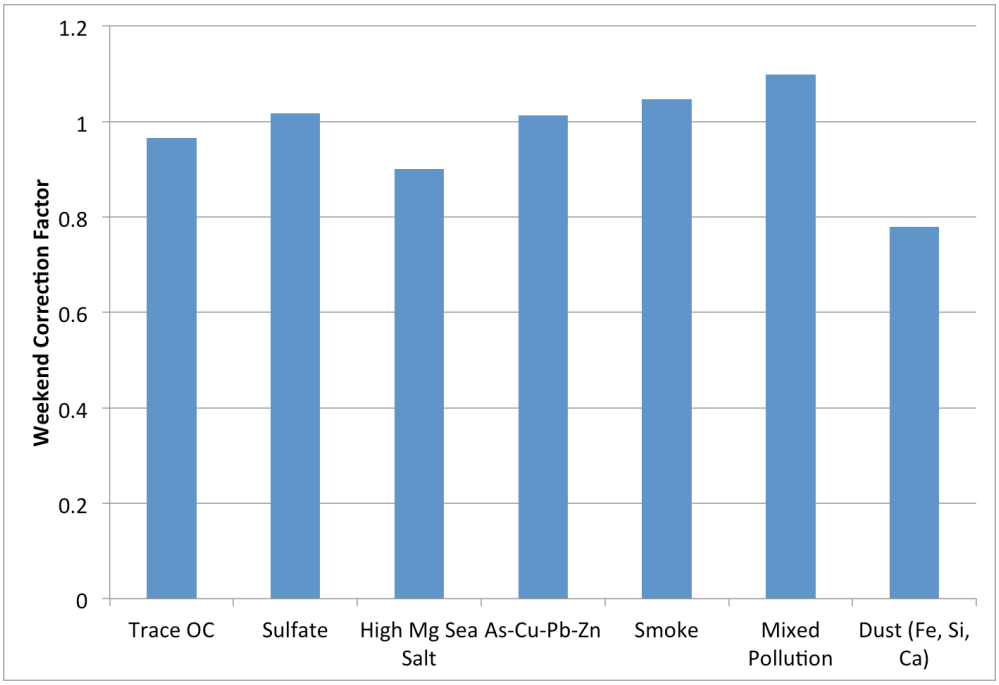

Figure 42. Weekend correction factors for the 7-factor extended parametric model fit. A value above 1 means that the factor is enhanced during that two-month period.

Contributions of Each Factor to $PM_{2.5}$ at BIBE

Figure 43 shows the relative contributions of the factors from the 7-factor extended parametric model fit averaged over all days between 2011 – 2014. The estimated contributions from dust (23 %) is basically the same as in the original 6-factor PMF analysis and in the 6-factor expanded parametric model fit. However, the contribution from smoke (12 %) is smaller (12 % instead of 17-18 %). "Sulfate" and "Mixed Pollution" together account for 42% of the $PM_{2.5}$, which is consistent with the "Sulfate" factor from the original PMF (43 %). The heavy metal "As-Cu-Pb-Zn" factor accounts for a significant amount of the mass (18 %), much larger than the equivalent factor in the original PMF fit (3 %).

The relative contributions of the factors to the days above the 80th percentile for $PM_{2.5}$ at BIBE (the "Top 20 %") are shown in Figure 44. The contributions from dust and smoke increase to 30 % and 16 %, respectively, while the contributions from the pollution factors decrease slightly to 38 %. The "As-Cu-Pb-Zn" factor is the biggest factor on days below the 20th percentile for $PM_{2.5}$ (34 %, Figure 45), while the two pollution factors decrease to 31 %.

Figure 46 to Figure 49 shows the time series of the different factor's absolute contributions to $PM_{2.5}$ at BIBE for each of the four years analyzed. As before, we can see strong episodic contributions of the "Smoke" factor in 2011, which was a severe drought year in Texas, as well as episodic high dust contributions due to dust storms. Contributions from the other factors appear to show less day-to-day and inter-annual variability.

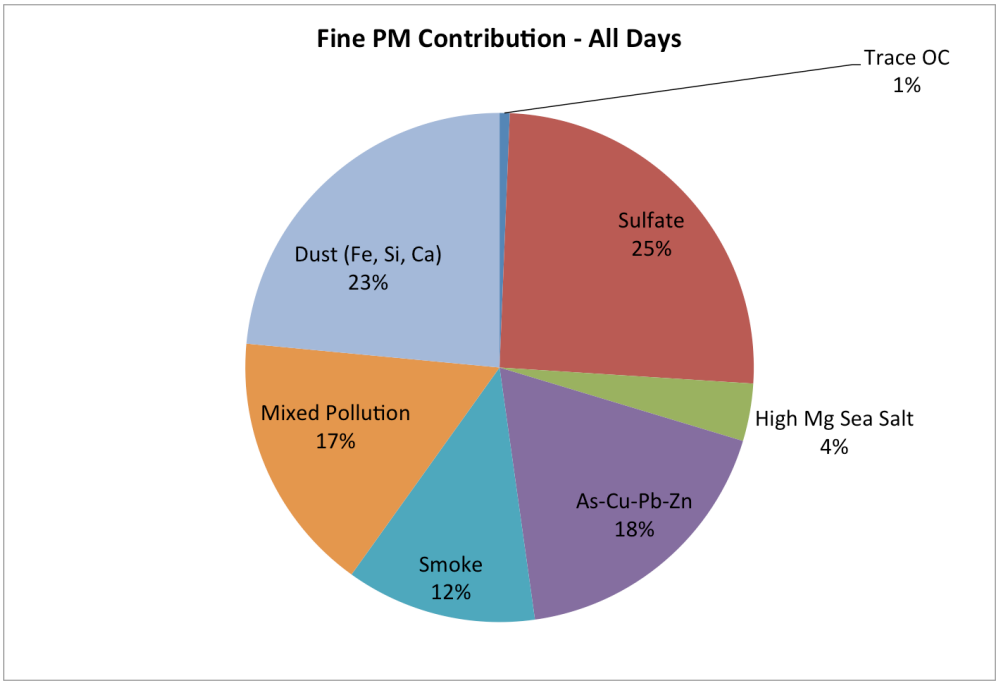

Figure 43. Relative contributions of the factors from the 7-factor extended parametric model fit averaged over all days at BIBE with IMPROVE aerosol measurements between 2011-2014.

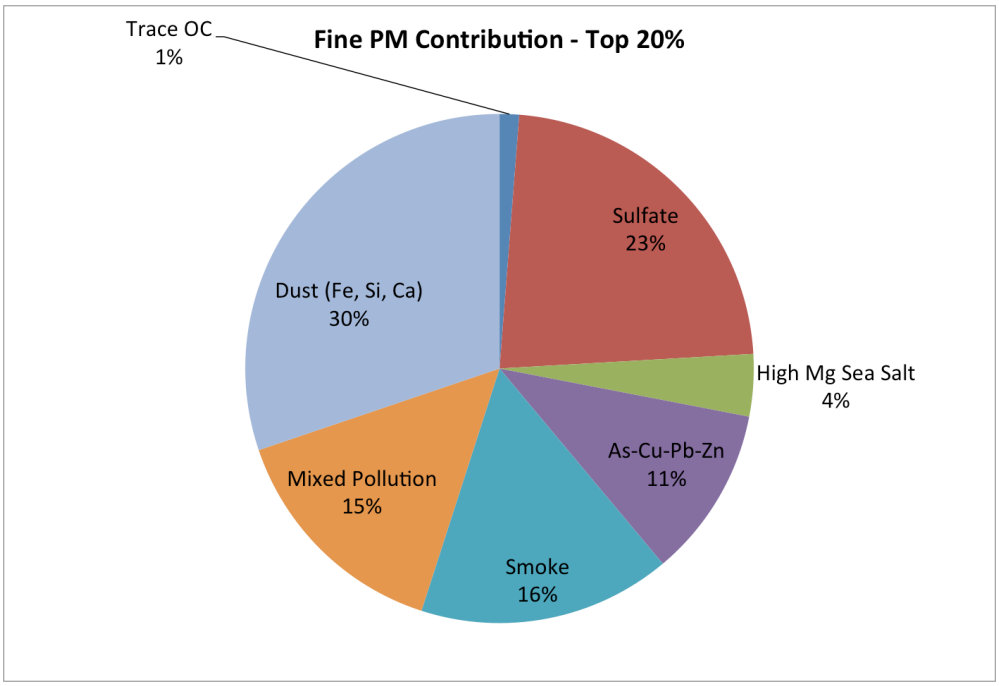

Figure 44. As in Figure 43, but only for the days above the 80th percentile for $PM_{2.5}$ at BIBE (the "Top 20%").

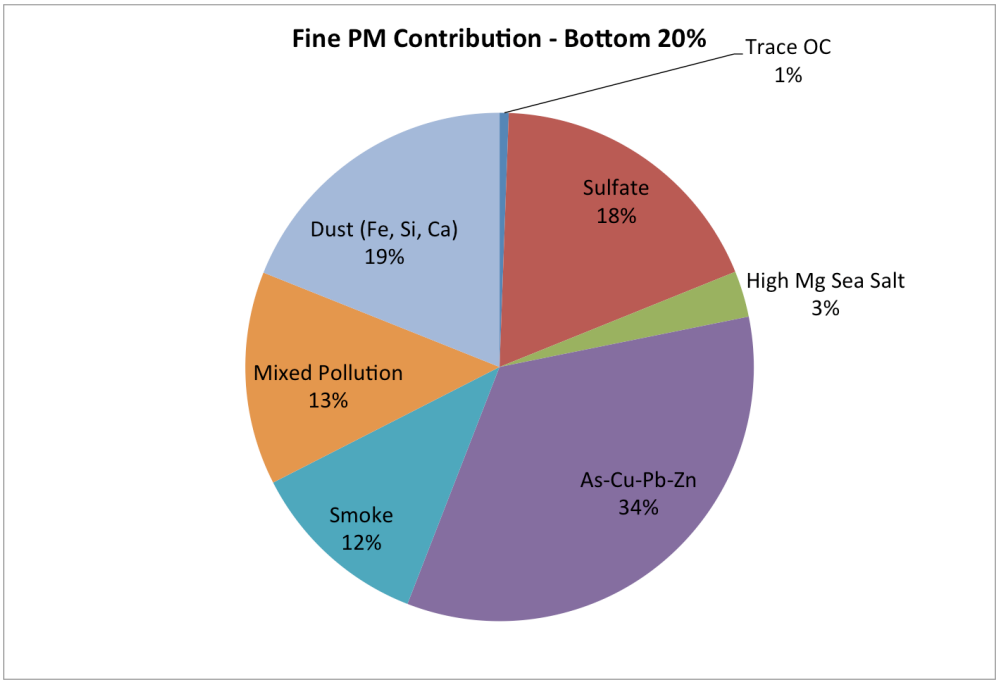

Figure 45. As in Figure 43, but only for the days below the 20th percentile for $PM_{2.5}$ at BIBE (the "Bottom 20%").

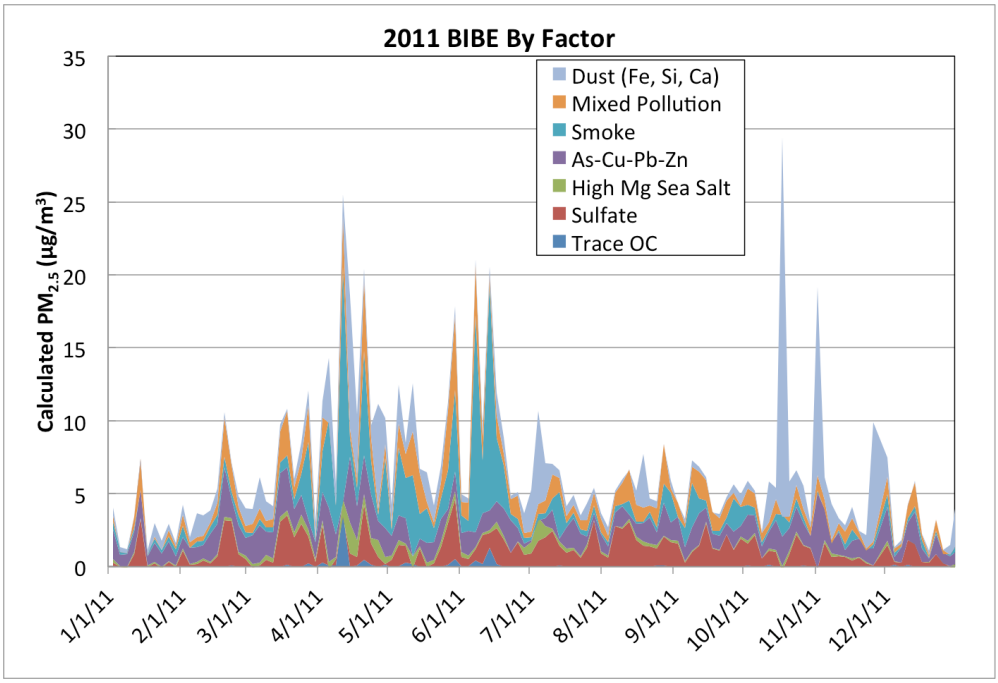

Figure 46. Time series of the absolute (µg/m^3) contributions of the factors from the 7-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2011.

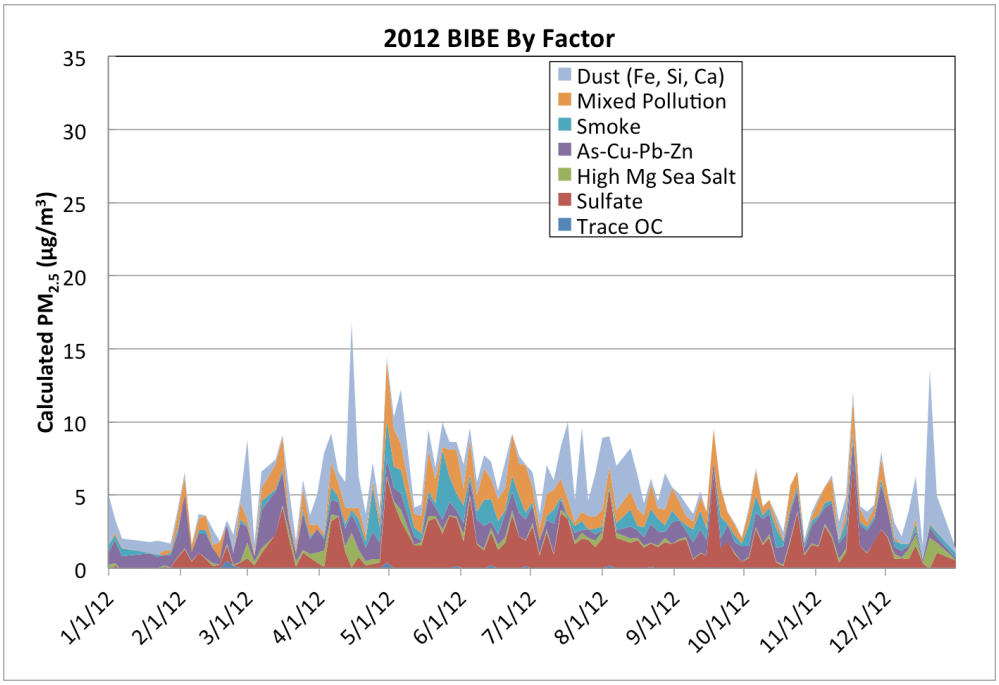

Figure 47. Time series of the absolute (µg/m^3) contributions of the factors from the 7-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2012.

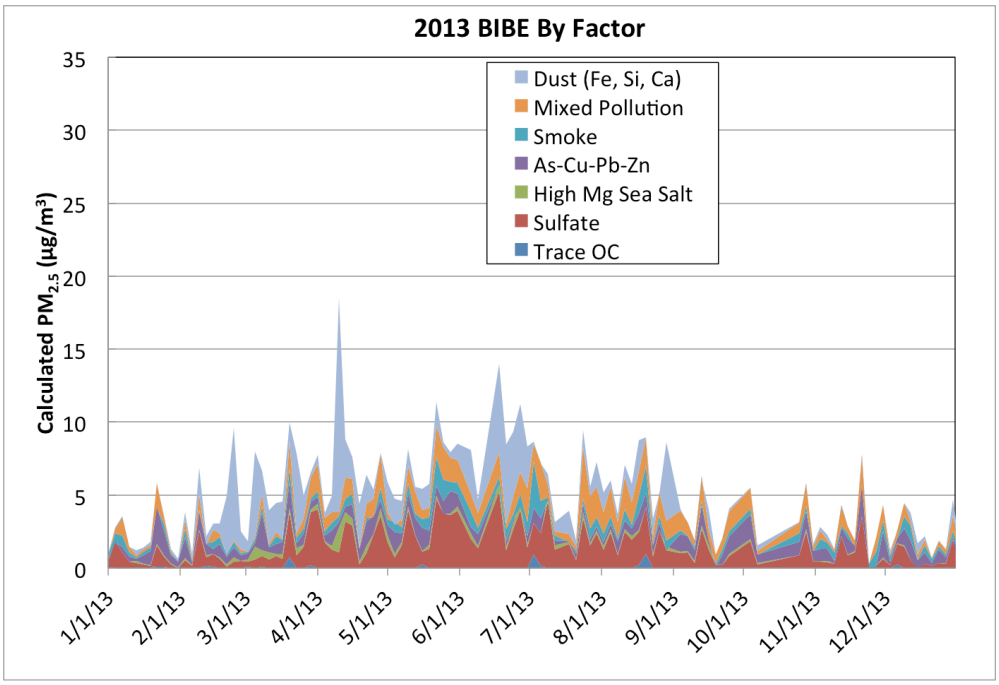

Figure 48. Time series of the absolute (µg/m^3) contributions of the factors from the 7-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2013.

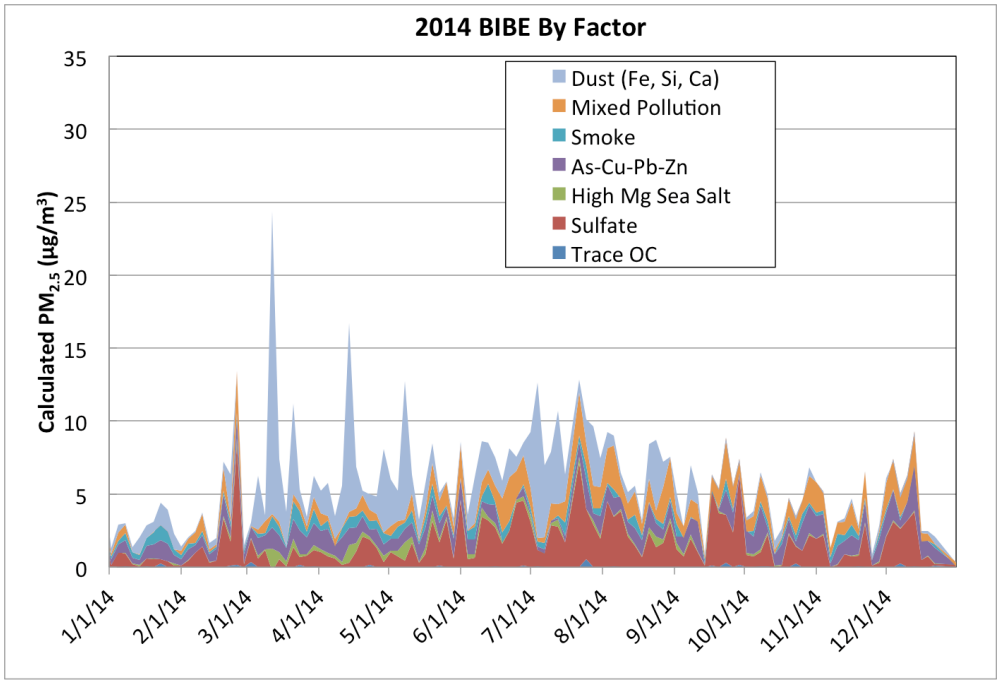

Figure 49. Time series of the absolute (µg/m^3) contributions of the factors from the 7-factor extended parametric model fit to $PM_{2.5}$ at BIBE for 2014.

3.3 Conclusions and Recommendations for Future Work

We find that the expanded parametric model included in the EPA PMF software is a reasonable way to add additional meteorological and other data (specifically, wind speed, wind direction, season, and day of week) to the PMF analysis. In our analysis, the best results were obtained when performing a 7-factor fit, as described in Section 3.2.2.

The advantage of the expanded parametric model is that the relationships between the factors and the wind data, as well as their seasonality and dependence on day of week, are direct outputs of the model. This can also help in identifying factors or determining when an identification based on the species profile may be incorrect. For example, some of the results from the 6-factor expanded model fit had odd dependence on wind speed given their identification, but in the 7-factor fit the dependencies on wind speed, as well as wind direction and seasonality, were consistent with our prior knowledge of the dependence of the sources on these variables.

However, the contribution estimates from the 7-factor expanded parametric model are not significantly different from our original 6-factor PMF fit. The estimated contributions from dust and anthropogenic pollution are nearly identical. The estimated contributions from smoke and heavy metals are different, but both analyses picked out the same factors, and the parametric model wasn't able to separate, say, pollution from the South (Mexico) from pollution from the East (Texas). In addition, using the expanded parametric model requires performing all of the analysis outside of the GUI of the EPA PMF software, and so could be more labor intensive. Thus we would not recommend using the expanded parametric model regularly, instead using it for cases where prior knowledge of the sources suggests that there should be a significant dependence of specific sources on wind direction, such as when two sources with similar species profiles are located near the measurement location but at different directions. A remote location like BIBE is thus probably not the best location for the use of the expanded parametric model.

Future work could explore other choices for setting up the indices for wind speed, etc., or adding additional variables to the parametric model (e.g., temperature). However, our results for BIBE do not suggest that further refinement of our approach would result in significantly different results.

4 Footprints

We used the Stochastic Time-Inverted Lagrangian Transport (STILT) model *[Lin et al., 2003]* with meteorological data from the North American Regional Reanalysis (NARR) to calculate the average daily surface footprints for each day in the study period. The footprint is a quantity representing the influence of upwind surface fluxes on concentrations measured at a receptor, defined as a specific location at a specific time, and is computed by counting the number of particles in a surface-influenced volume, defined as lower half of the planetary boundary layer, and the time spent in that region. The footprint method provides source attribution information similar to that of the trajectory residence time method used by *Schichtel et al. [2006]* to associate air mass transport patterns with particulate sulfur concentrations during BRAVO *[Pitchford et al., 2004]*. However, since footprints only include particles that pass near the surface, they are a more accurate representation of the influence of surface emissions from a given location than trajectory residence times that do not explicitly take into account the height of the trajectory as it passes over a surface location. Furthermore, footprints are quantitative and can be used to calculate the contribution to the concentration of a species at a downwind receptor if surface fluxes are known and chemical production and loss along the trajectory is negligible.

For this study 10-day footprints in 0.5° x 0.5° grid cells were generated for receptors located at the BIBE IMPROVE site for each day. We examined footprints for individual days when specific factors are high to gain a better understanding of the transport patterns influencing the speciated particulate matter concentrations measured at BIBE. For example, Figures 50 to 52 show average footprints for 3 days with differing source contribution factor profiles. On January 15, 2011 (Figure 50) the sulfate source contribution factor was relatively high at 6.58 µg m^{-3} while most other factor contributions were low, and the footprint analysis indicated that most of the surface influence was from Mexico. In contrast, on June 2, 2011 (Figure 51) the sulfate factor was only 0.61 µg m^{-3} while the smoke factor was moderately elevated at 2.78 µg m^{-3}. For this day the airmass transport showed little surface influence from North America but perhaps some inter-continental transport of smoky air from biomass burning. On August 28, 2011, the sulfate factor was high at 5.32 µg m^{-3} and smoke was moderate at 1.6 µg m^{-3}. On this day the surface influence was from eastern Texas, which includes the cities of Houston, Dallas and Austin, and the south central US (Figure 52). This transport pattern is consistent with high anthropogenic influence mixed with some fires.

Then we attempted to use the footprints to calculate a quantity analogous to the Potential Source Contribution Function (PSFC) described by Phillip Hopke in a presentation to the California Air Resources Board.[1] Hopke defines the PSFC as

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \qquad (5)$$

where m_{ij} is the number of back trajectory endpoints that pass through grid point *i,j* for a sampling interval when the source contribution factor value is greater than the 60th percentile value of that particular source contribution factor and n_{ij} is the total number of back trajectory endpoints that pass through the same grid point during the entire sampling period. The resulting PSCF quantity is a probability field representing the likely source locations of the material that results in high

[1] *http://www.arb.ca.gov/carbis/research/seminars/hopke/carb_presentation.pdf*

measured concentrations at the receptor site. For our analogous PSFC we replaced the trajectory end point counts of m_{ij} and n_{ij} with the total footprints for the days when the particular source contribution factor is greater than its 60th percentile value and for all the days respectively.

An example of our analogous PSCF footprint-based quantity is shown in Figure 53. A problem we discovered with this approach is that the PSFC increases far away from the receptor location due to the fact that the total footprint values in the denominator of Equation 5 tend to be approach zero away from the receptor. The footprint calculation only considers particles near the surface and it becomes less and less likely that a particle will be near the surface the further away it gets from a near surface receptor and hence the footprint values tend toward zero at longer distances away from the receptor. The trajectory-based approach of Hopke is spared this problem because it does not consider the height of a trajectory point only its geographic location in a horizontal grid. Therefore, for the remainder of this work we utilized only average footprints as shown in Figures 50 to 52 but for the entire study period.

The average footprint field for all the days from 2011 – 2014 with $PM_{2.5}$ concentrations in the upper 20 % indicates that the main source contribution region was in north central Mexico (Figure 54). Relatively strong source contribution regions also included southwest Texas and eastern Texas. For days with $PM_{2.5}$ concentrations in the lower 20 % north central Mexico and central Texas were strong source regions but the overall footprint magnitude was lower than for the high $PM_{2.5}$ days (Figure 55). Furthermore, there were lower contributions from eastern Texas but considerable contributions from a region extending across the Gulf of Mexico to the eastern edge of the domain over the Caribbean Sea.

The average footprints for each source contribution factor for the factors identified by the baseline PMF runs described in Section 2 are shown for days with both $PM_{2.5}$ concentrations in the upper 20 % and factor contributions greater than the 60th percentile value in Figure 56. There were too few days with both $PM_{2.5}$ concentrations in the lower 20% and factor contributions greater than the 60th percentile to produce meaningful average footprint fields and these are not shown. From these plots there were no clear differences in the major source contribution regions. For all plots the strongest contribution region seemed to be from north central Mexico just south and southwest of BIBE. This suggests a strong prevailing wind pattern with little synoptic variability. This is a likely hypothesis given the location of BIBE in the sub-tropical latitude belt and a significant distance from major water bodies.

Despite the consistency in the location of the primary source region there were some general differences in the secondary regions and the overall magnitudes of the footprints. The Sulfate factor footprints showed the strongest contributions from eastern Texas. The Smoke, Sulfate, Dust (Ca, Fe, Si) and Nitrate factors all had significant contributions from western and central Texas. Additionally, the Sea Salt / Nitrate mixture, Dust (Ca, Fe, Si) and Nitrate factors had contributions from the southeast of BIBE extending across the southern Gulf of Mexico into the Caribbean. Overall the footprint analysis indicated a strong source contribution region for all factors from north central Mexico but secondary strong contribution regions for Smoke, Dust (Ca, Fe, Si) and Sulfate and Nitrate pollution from central and western Texas with an additional region for Sulfate and Smoke from eastern Texas.

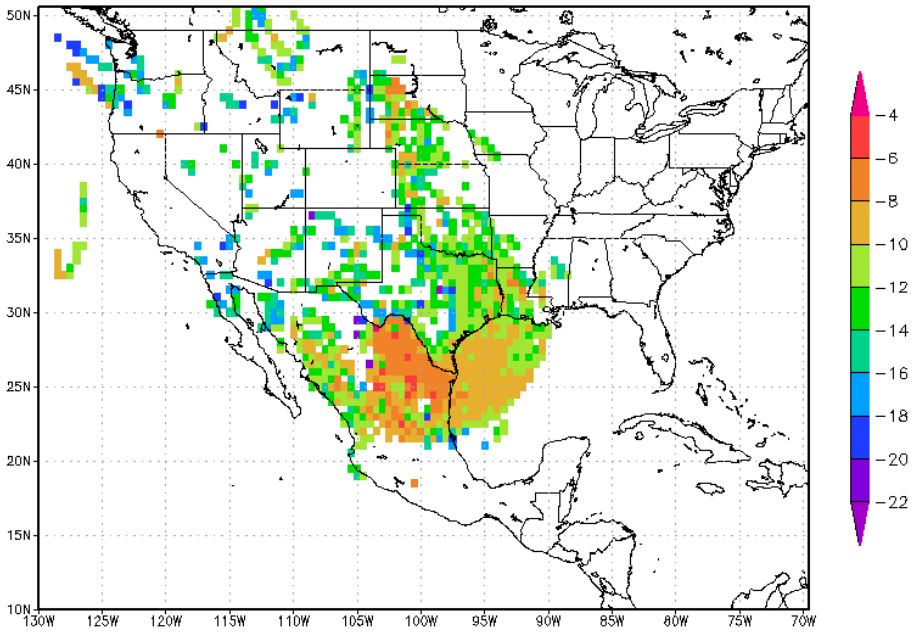

Figure 50. Average footprints for January 15, 2011 for BIBE. The footprints expressed as ln(ppmv/micromol/m^2/s) represent the surface influence experienced by air parcels arriving at the BIBE receptor.

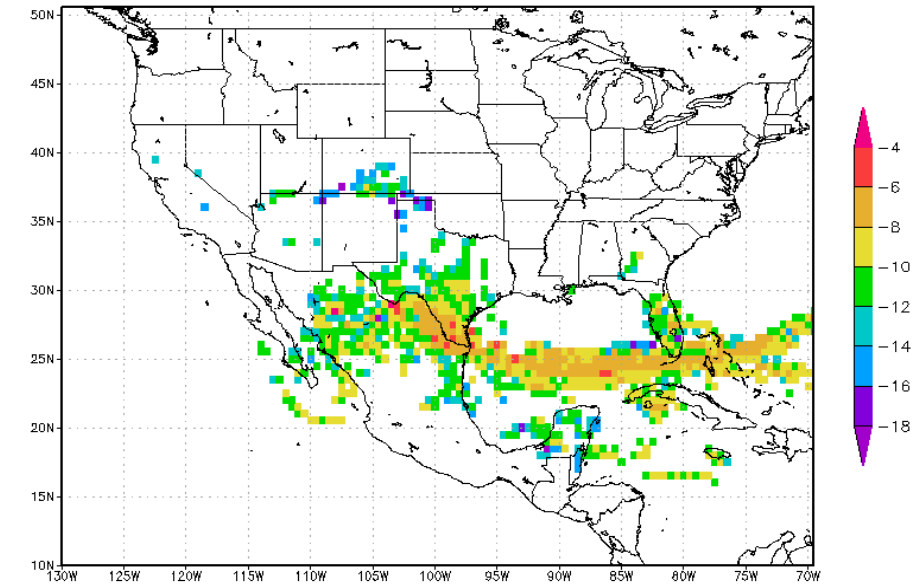

Figure 51. Same as Figure 50 but for June 2, 2011.

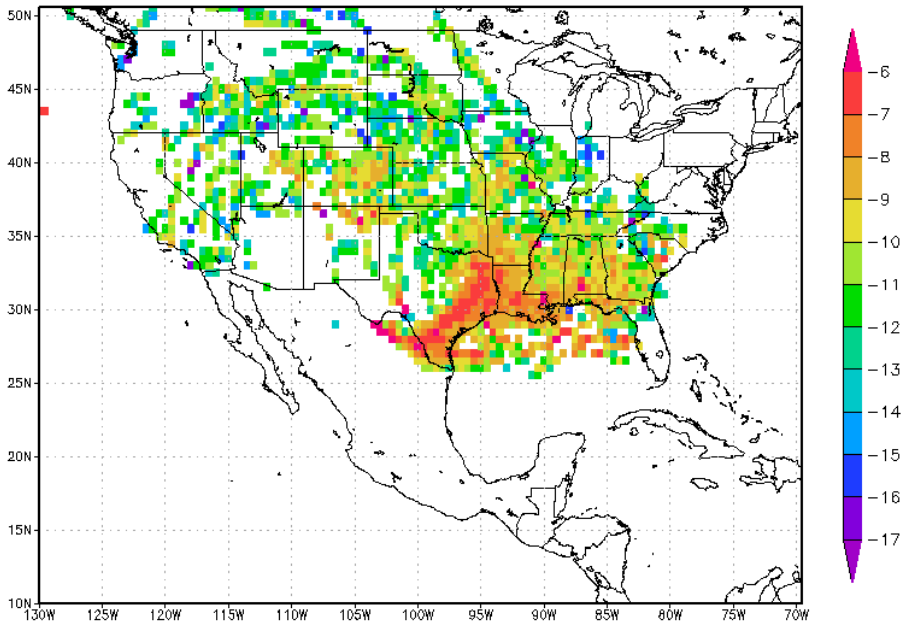

Figure 52. Same as Figure 50 but for August 28, 2011.

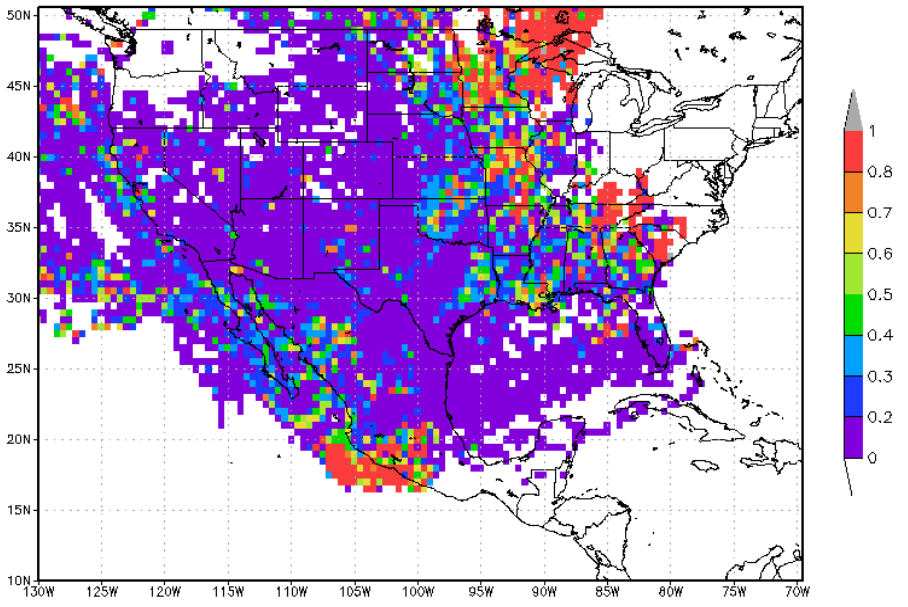

Figure 53. PSCF based on STILT footprints for days when the factor 3 "Sulfate" contribution is greater than the 60th percentile value and the $PM_{2.5}$ concentration is in the upper 20 %.

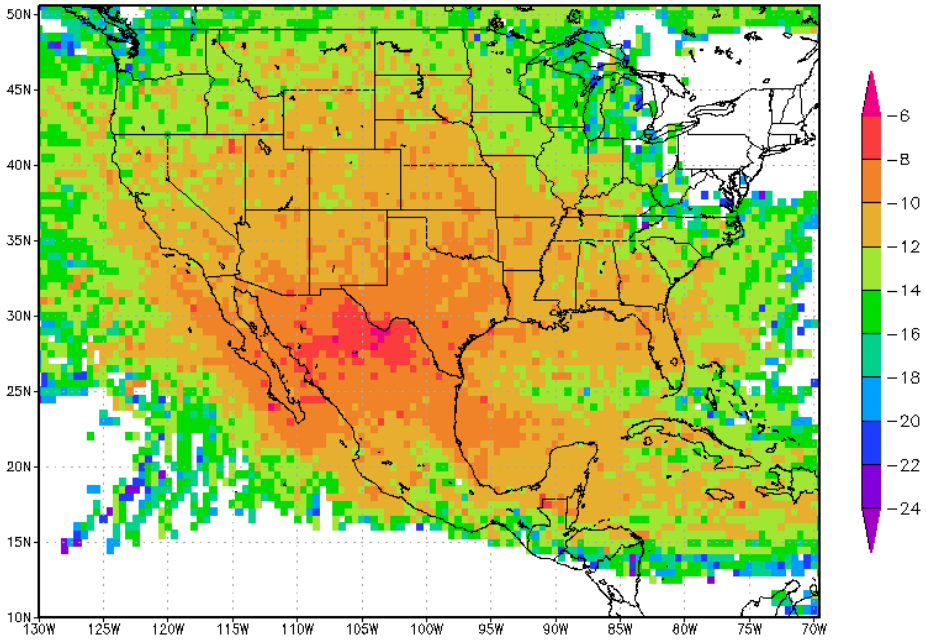

Figure 54. Average 2011 – 2014 footprints for all the days with greatest 20 % $PM_{2.5}$ concentrations measured in BIBE1. The footprints expressed as ln(ppmv/micromol/m^2/s) represent the surface influence experienced by air parcels arriving at the BIBE receptor.

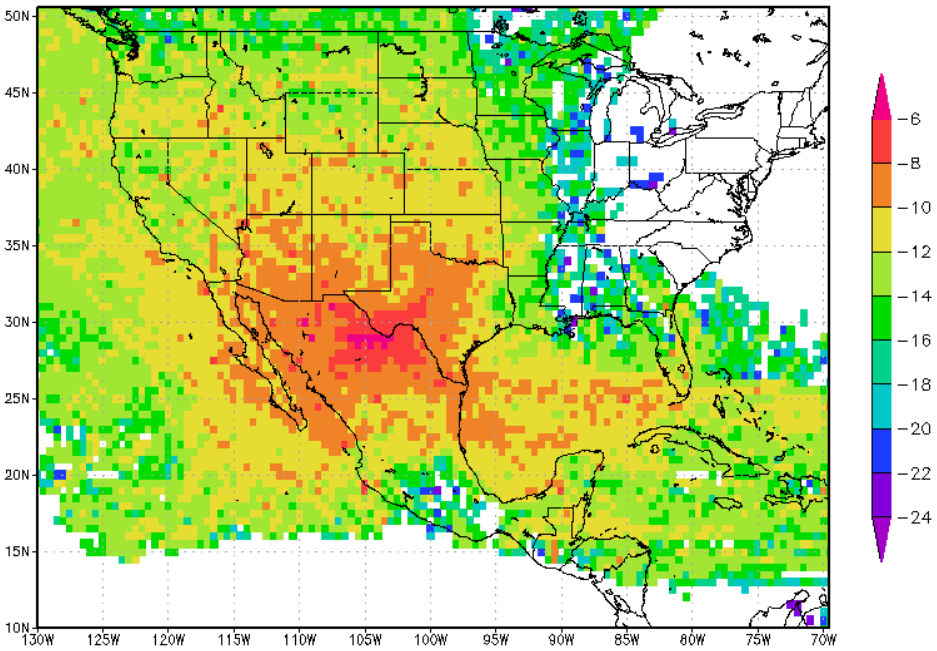

Figure 55. Average 2011 – 2014 footprints for all the days with lowest 20 % $PM_{2.5}$ concentrations measured in BIBE1. The footprints expressed as ln(ppmv/micromol/m^2/s) represent the surface influence experienced by air parcels arriving at the BIBE receptor.

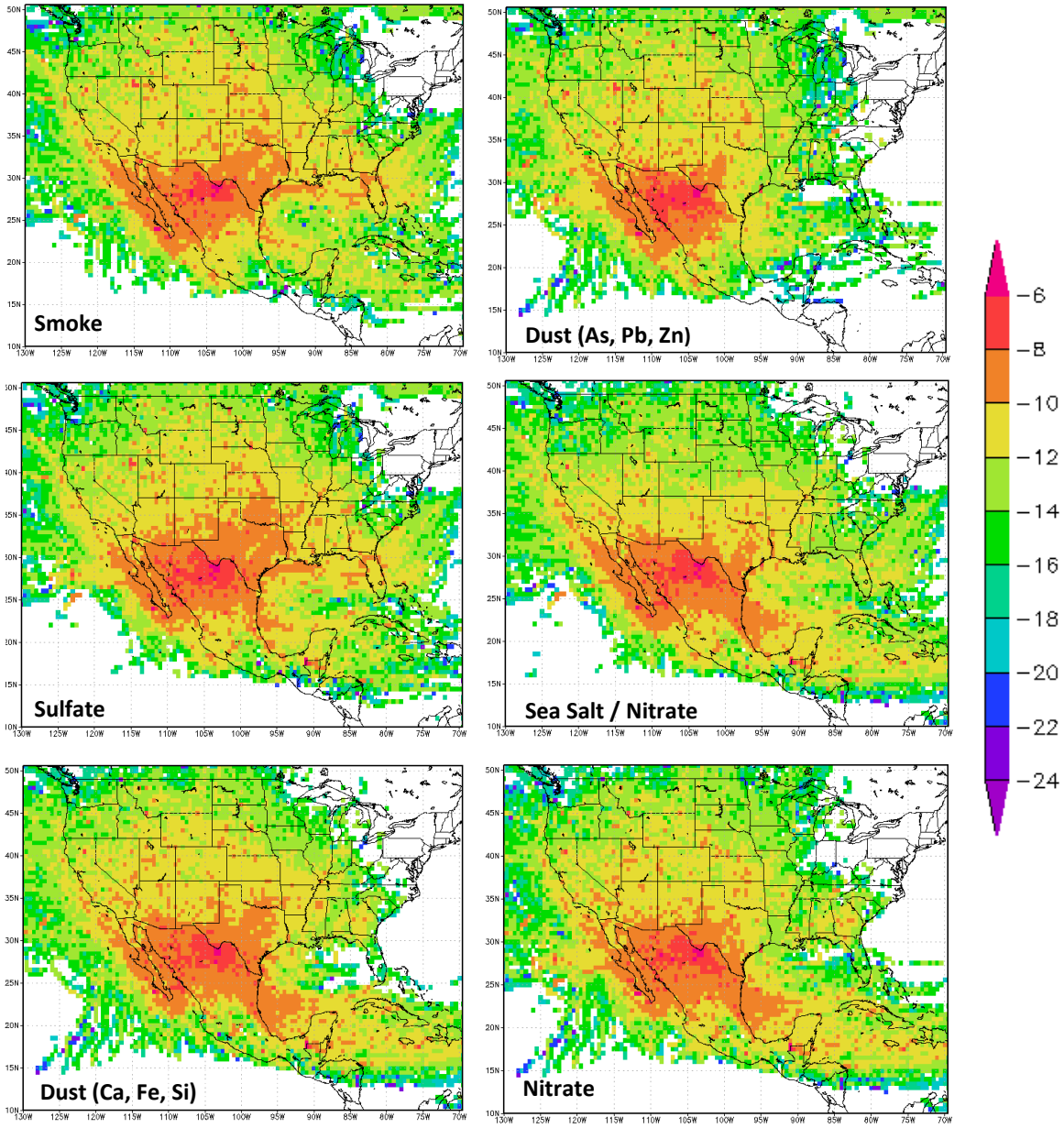

Figure 56. Average 2011 – 2014 factor footprints for all the days with greatest 20 % $PM_{2.5}$ concentrations measured in BIBE1and factor contributions greater than the 60th percentile value. The factors are those identified from the 6-factor baseline PMF run described in Section 2. The footprints are expressed as ln(ppmv/micromol/m^2/s) represent the surface influence experienced by air parcels arriving at the BIBE receptor.

5 Summary and Recommendations for Future Study

This project investigated the sources of fine particulate matter ($PM_{2.5}$) in Big Bend National Park (BIBE) during 2011 – 2014 using speciated aerosol measurements from the IMPROVE monitoring network. The investigation utilized the Positive Matrix Factorization (PMF) technique to identify the major source species and their relative contributions to the $PM_{2.5}$ concentrations. An initial 6-factor baseline run using the GUI driven EPA 5.0 PMF tool was performed on the speciated aerosol profiles and the results of this run were compared with a previous study conducted by the Causes of Haze Assessment (COHA) program during the 2000 – 2004 time frame. We also performed a qualitative examination of the relationship between meteorological and time variables and the source contribution factors identified by the baseline run and attempted to incorporate these variables into the GUI PMF tool but found that this was not feasible. However, we found that meteorological and time variables could be rigorously incorporated into the PMF analysis using the Expanded Parametric Model included with EPA PMF 5.0 software but not accessible from the GUI. Finally, we also performed an investigation into the locations of the various species source contributions using footprints generated by the STILT Lagrangian transport model. Our investigation produced several key findings as summarized below.

The major contributing factors to the $PM_{2.5}$ concentrations in BIBE during 2011 – 2014 were secondary sulfate pollution, dust (mostly mineral or soil dust), and biomass burning smoke while other factors such as secondary nitrate pollution, sea salt, and heavy metal dust were of lesser importance. This finding was generally consistent with the COHA analysis of 2000 – 2004 that identified the same major contribution factors. However, an important difference with the earlier assessment was that for the days in 2011 – 2014 with the highest 20 % of $PM_{2.5}$ concentrations the proportional contribution of sulfate pollution decreased substantially and was less than the combination of contributions from natural sources such as mineral dust and biomass burning smoke. This result suggests a decreasing role of sulfate pollution in contributing to high $PM_{2.5}$ concentrations in BIBE.

Local meteorology influenced the aerosol concentrations as days with the highest 20 % of $PM_{2.5}$ were warmer and more humid than the days with the lowest 20 % of $PM_{2.5}$. Sulfate, smoke, sea salt and nitrate factors were most strongly affected. On the other hand, local wind variability was not highly correlated with any of the source factor contributions or total $PM_{2.5}$ concentrations. The factor contributions showed seasonal variability with smoke peaking in spring and dust in the fall, while sulfate pollution seemed to peak on Tuesdays and Saturdays possibly indicating a time lag due to transport time from distant emissions sources. We found that the Expanded Parametric Model is a reasonable way to add additional meteorological and other data (specifically, wind speed, wind direction, season, and day of week) to the PMF analysis. In our analysis, the best results were obtained when performing a 7-factor fit; however, the contribution estimates from the 7-factor Expanded Parametric Model are not significantly different from our original baseline 6-factor PMF fit.

Our footprint analysis was able to identify plausible source regions of secondary sulfate pollution for individual case days but could not clearly delineate different source regions corresponding to different source factors, as the average footprints indicated that the strongest source region for all factors was from south and west of BIBE in north central Mexico. However, strong secondary source contribution regions were identified for sulfate, smoke, dust and nitrate in western and central Texas and for smoke and sulfate pollution in eastern Texas.

Based on this investigation, we have several recommendations concerning future study of fine particulate aerosols in BIBE. The baseline PMF model run through the GUI seems to be an adequate tool for assessing the source contributions to $PM_{2.5}$ concentrations in BIBE. Though the Expanded Parametric Model is a reasonable way to add additional meteorological and other data to the PMF analysis, it provided little additional information to our study. Furthermore, the Expanded Parametric Model requires performing all of the analysis outside of the GUI of the EPA PMF tool, and so could be more labor intensive. Thus we would recommend using the Expanded Parametric Model only for cases where prior knowledge of the sources suggests that there should be a significant dependence of specific sources on wind direction, such as when two sources with similar species profiles are located near the measurement location but at different directions. A remote location like BIBE is thus probably not the best location for the use of the expanded parametric model.

To better assess the regions associated with the source factors it may be helpful to run the STILT model or any other Lagrangian transport model with higher resolution meteorology inputs. For this investigation we used the meteorological inputs NARR available on a 32-km grid every 3 hours and this resolution may not have been adequate to resolve finer-scale meteorological features that may have influenced the footprints. Therefore, we recommend testing the impact of meteorological resolution on source location attribution for future studies. In addition, a finer footprint grid on the order of a few kilometers may help highlight specific facilities producing pollution within larger regions.

6 References

Chan, Y.-c., O. Hawas, D. Hawker, P. Vowles, D. D. Cohen, E. Stelcer, R. Simpson, G. Golding, and E. Christensen, 2011: Using multiple type composition data and wind data in PMF analysis to apportion and locate sources of air pollutants, *Atmos. Environ.*, 45, 439-449.

Green, M., 2006: Causes of Haze Assessment, online available report available at http://coha.dri.edu/index.html.

Lin, J. C., C. Gerbig, S. C. Wofsy, A. E. Andrews, B. C. Daube, K. J. Davis, and A. Grainger, 2003: A near-field tool for simulating the upstream influence of atmospheric observations: The stochastic time-inverted Lagrangian transport model (STILT), *J. Geophys. Res., 108(D16)*, 4493, 10.1029/2002JD003 161.

Norris, G., R. Duvall, S. Brown, and S. Bai, 2014: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, US EPA/600/R-14/108, 136 pp, available at http://www.epa.gov/sites/production/files/2015-02/documents/pmf_5.0_user_guide.pdf.

Paatero, P., 1999: The multilinear engine – a table-driven least squares program for solving multilinear problems, including the n-way parallel factor analysis model, *J. Comput. Graphical Stat.*, 8, 854-888.

Paatero, P., 2009: Extended factor analysis, enhanced with error estimation of composition factors and with different rotational tools, including PMF2-style FPEAK rotations, internal memo prepared for EPA, June 17, 2009, 38 pp.

Paatero, P., and P. K. Hopke, 2002: Utilizing wind direction and wind speed as independent variables in multilinear receptor model studies, *Chemom. Intell. Lab. Syst.,* 60, 25-41.

Pitchford, M. L., I. Tombach, M. Barna, K. A. Gebhart, M. C. Green, E. Knipping, N. Kumar, W. C. Malm, B. Pun, B. A. Schichtel, and C. Seigneur, 2004: Big Bend Regional Aerosol and Visibility Observational Study Final Report, September 2004, 310 pp., available at http://vista.cira.colostate.edu/improve/Studies/BRAVO/reports/FinalReport/bravofinalreport.htm.

Schichtel, B. A., K. A. Gebhart, M. G. Barna, and W. C. Malm, 2006: Association of airmass transport patterns and particulate sulfur concentrations at Big Bend National Park, Texas, Atmos. Environ., 40, 992-1006.

Appendix A: Running the Extended Parametric Model

A.1 Setup

In these instructions, "root directory" is the directory that your copy of the EPA PMF software is installed in, most likely C:\Program Files (x86)\EPA PMF\EPA PMF 5.0\ . We also recommend that you read the following technical documents, which are provided with this report:

- *Param_info_v5.txt* – a README file that describes the implementation of the extended parametric model in ME2 and the EPA PMF software.
- *Math_Extended_FA_v41.doc* – A Word file documenting the mathematics behind the extended parametric model.

1. First perform a PMF run using the EPA PMF GUI. This will allow you to use the GUI to set the PMF configuration (which variables are included, which are assigned "weak" or "bad", etc.), and you will need the inputs prepared by the GUI for the expanded parametric model.
2. You'll need two files from the root directory:
   2.1. iniParams_copy.txt – This is a copy of the input control parameter file for ME2 that the GUI made.
   2.2. PMFData.txt – This is the input data file for ME2 (CSV format), which was prepared by the GUI. The format is a (row = day, column = species) concentration matrix immediately followed by a (row = day,column = species) uncertainty matrix. I think that the species are in the same order as the original input file EXCEPT that any species marked "BAD" in the GUI have been removed. It's probably best to figure out the species to column mapping here, because you'll need it later.

   NOTE: We have found that the above files are only saved after the PMF run on some Windows machines, but not on others. We are not sure what is the difference in the machines that causes this.
3. You'll need to make the data file for the additional variables (the "parametric factors") and name it "pf_inds.txt". Below are the first few lines of the file used in our analysis:

```
4   16   6
0    1   1
3   11   1   0
2   11   1   0
3   11   1   1
```

   Line 1 gives the maximum index number for each of the parametric factors that isn't a simple binary variable (like weekend versus weekday). In our case, these are the maximum indices for the wind speed (4), wind direction (16), and season (6).

   Line 2 gives the "circular indicator" for the variables. This is 1 for periodic variables (like wind direction and season) and 0 for variables like wind speed.

   Afterwards, you must have exactly one line for each date included in PMFdata.txt and each line must have the values for all of the parametric variables. The non-binary variables must go first in the same order as on Lines 1 and 2. For the non-binary variables, a '0' indicates missing data. The binary (yes/no) variables, like weekend/weekday, go afterward, with a "1" meaning

"yes" (in our case, that the date is on the weekend) and a "0" meaning "no" (the date is on a weekday).

4. Copy iniParams_copy.txt to a separate file (e.g., iniParams_safe.txt) to prevent it being overwritten.
5. To use the extended parametric PMF model, make the following changes to iniParams_safe.txt:
   5.1. On Line 11, change the value for "modifcode" to "1"
   5.2. On Line 33, change:
      5.2.1. "numpf" to the number of additional non-binary variables (i.e., "parametric factors" that have maximum indices greater than 2). In our study, these are wind speed, wind direction, and season, so "numpf" is set to 3.
      5.2.2. "numynpf" to the number of binary (yes/no) additional variables. In our study, we only have the weekday/weekend variable, so this is set to 1.
      5.2.3. "maxpfdim" is the maximum dimension for the additional variables, which for our study is 16 (the number of wind direction indices).
6. Copy iniParams_safe.txt to iniParams.txt
7. Open a DOS prompt (Command Prompt) and change to the root directory,
8. Run ME2 with a command like "me2gfP4_1345c4.exe PMF_bs_6f8xx_sealed_GUI.ini", where the first is the ME2 executable file name in the root directory and the second is the PMF*.ini file in this directory.
9. Copy all the output files produced to a different directory.

A.2 Output Format

1. *PMF_ab_base.txt:* This is the best-documented output file.
   1.1. Top of file repeats the input control variables from iniParams.txt.
   1.2. Then the file has output for each of the PMF runs (default is 20 runs), including:
      1.2.1. The goodness-of-fit Q statistic: The run with the lowest value of Q(robust) should be chosen for further analysis.
      1.2.2. AA matrix – this is the normalized "G" matrix, with a row (line) for each day and a column for each factor. So each row is the normalized contribution of each factor on that day. The matrix is normalized such that the average of each column is 1. This can be converted into physical units (e.g., $\mu g/m^3$ contribution of each factor to fine particulate matter) by multiplying each column by the corresponding value for "MF" from the BB matrix below.
      1.2.3. BB matrix – this is the normalized "F" matrix, with a row (line) for each species and a column for each factor. Each column is the species profile for the given factor, and you can normalize each row to calculate the percentage of a given species in each factor. .The row for the "MF" variable also gives the scale factors you need to convert the normalized factor contributions in the AA matrix into physical concentrations.
      1.2.4. CC matrix – A single row matrix with all values set to 1.0.
      1.2.5. PF matrices – these are the matrices for the four additional variables – (1) wind speed, (2) wind direction (3) season and (4) weekend versus weekday.
   1.3. At the bottom of the file, there are additional summary statistics for the 20 runs.

2. *PMF_ab_base.dat:* This has the same information as the *.txt file, but without the extra documentation. The output is for each of the 20 runs, in this format:
   2.1. AA matrix (one line per row/day)
   2.2. Blank line
   2.3. BB matrix (one line per row/species)
   2.4. Blank line
   2.5. CC matrix
   2.6. Blank Line
   2.7. "Stacked" PF matrix – This is a set of four matrices, printed one after another (no blank line), that correspond to the matrices for the four additional variables – (1) wind speed, (2) wind direction (3) season and (4) weekend versus weekday. Each matrix has dimensions maxpfdim (16 in our runs) by the number of factors fitted. For variables with indices less than maxpfdim, the extra values are set as 0.

3. *PMF_report.txt:* Space-delimited file with total Q statistics and other summary parameters for each run.

A.3 Analysis

For our analysis, we copied the AA, BB, and PF matrices for the best-fit run (lowest Q(robust) value) from PMF_ab_base.txt into an Excel spreadsheet for plotting. Our example input and output files, as well as the Excel spreadsheet used for plotting, are delivered with this final report.